Cours Concis de Mathématiques

Pierre Guillot
IRMA
7 rue René Descartes
67084 Strasbourg
guillot@math.unistra.fr

Avant-propos

> – Deux ans ! dit Dantès, vous croyez que je pourrais apprendre toutes ces choses en deux ans ?
> – Dans leur application, non ; dans leurs principes, oui : apprendre n'est pas savoir ; il y a les sachants et les savants : c'est la mémoire qui fait les uns, c'est la philosophie qui fait les autres.
> [. . .]
> Dantès avait une mémoire prodigieuse, une facilité de conception extrême : la disposition mathématique de son esprit le rendait apte à tout comprendre par le calcul, tandis que la poésie du marin corrigeait tout ce que pouvait avoir de trop matériel la démonstration réduite à la sécheresse des chiffres ou à la rectitude des lignes.
>
> Alexandre Dumas, *Le comte de Monte-Cristo*

Edmond Dantès, le héros du roman d'Alexandre Dumas, dispose de deux ans pour apprendre les mathématiques, la physique, l'histoire, et « trois ou quatre langues vivantes », à l'insu des gardiens de sa prison. Votre tâche est plus abordable : si en un an vous venez à bout du présent ouvrage, vous aurez couvert les programmes d'analyse et d'algèbre de première année post-baccalauréat.

Certains étudiants se contenteront, selon leur filière, d'une sélection de quelques chapitres. Par exemple l'auteur a enseigné à des élèves chimistes, parcourant au premier semestre les chapitres 3, 4, 6, et le tout début des 8, 9 et 10, alors que le deuxième semestre, dédié à l'algèbre, s'est construit autour des chapitres 12, 13, 14, 15 et 16 (quelques morceaux choisis de la partie « Préliminaires » étant disséminés au fur et à mesure des besoins). À l'inverse, d'autres élèves amenés à étudier plus de mathématiques en première année, comme par exemple ceux en classes préparatoires, exploreront tous les chapitres et détailleront même les encadrés « hors programme ». Les enseignants infléchiront, par le choix des exercices supplémentaires qu'ils voudront bien proposer, votre course à travers l'analyse et l'algèbre.

Les encadrés

Au gré du texte vous verrez quelques encadrés, ayant l'apparence de celui-ci. Ils contiennent des compléments qui sont clairement à considérer comme du « hors programme ». Leur lecture demande plus d'attention.
Si un thème vous intéresse, vous êtes invité à faire quelques recherches, à commencer par Wikipedia et autres, puis à questionner vos enseignants.

Le parti pris de ce cours

L'idée de ce « cours concis » est de proposer, aux étudiants comme aux enseignants, un cours de mathématiques essentiellement identique à ce que l'auteur produirait au tableau. (On a bien sûr incorporé dans le texte les nombreuses remarques orales qui accompagneraient un tel cours.) Afin de tendre vers cet objectif, on a déshabillé ce livre du caractère encyclopédique dont est empreinte la majorité des ouvrages de mathématiques à l'attention des élèves de première année.

Vous pourrez d'ailleurs, si telle est votre préférence, faire une utilisation complémentaire de ce cours et d'un ou plusieurs ouvrages à vocation exhaustive.

Les exercices

Tout au long des chapitres, vous êtes invités à résoudre un certain nombre d'exercices. Toutefois, les énoncés de ceux-ci ne sont pas inclus dans le livre. Ils sont à retrouver sur le site Internet « exo7 », qui (au moment où cet avant-propos est rédigé, en mai 2013) se trouve à l'adresse suivante :

http://exo7.emath.fr/search.php

Vous y trouverez un très grand nombre d'exercices supplémentaires, dont une bonne partie comprend une correction, et parfois même une correction *filmée*. Sur la page personnelle de l'auteur, que vous trouverez aisément à l'aide d'un moteur de recherche, vous pourrez également télécharger les exercices au format PDF, ainsi qu'une version gratuite de ce livre.

Dans le texte, les numéros des exercices sont donnés dans la marge. Ils peuvent être reportés directement dans le « cadre de saisie » d'exo7.

Les notes dans la marge ont cet aspect.

Précisons une chose. La sélection des exercices que nous avons opérée est orientée vers la compréhension des concepts les plus basiques ; à de très rares exceptions près, les problèmes à résoudre ne font pas appel à des astuces. Le but est de vous assurer de votre bonne progression dans le cours, avec ce baromètre simple : vous devez être capable de faire *tous* les exercices ou presque.

Les exercices plus « astucieux », qui peuvent très bien être au goût de l'auteur par ailleurs, vous seront proposés par vos enseignants.

Les deux lectures

Un certains nombre de découpages ont été faits, dans le but de séparer les choses les plus faciles des choses les plus ardues. Les encadrés, on l'a dit, contiennent des informations officiellement « hors programme ». Mais vous constaterez également que la plupart des chapitres se partagent entre une « première lecture » et une « deuxième lecture ». La deuxième lecture renferme parfois les démonstrations de certains résultats annoncés dans la première ; parfois on y aborde d'autres concepts plus avancés.

Précisons qu'on attend d'un élève de première année en filière mathématique à l'université qu'il connaisse le contenu des « deuxièmes lectures » (et les enseignants en deuxième année supposeront que c'est le cas). À l'extrême inverse, nous recommandons aux étudiants « pressés » (pour désigner ainsi ceux qui entament le travail à quelques jours de l'examen) de commencer par les premières lectures de tous les chapitres (plutôt que de lire quelques chapitres en entier).

Notons que l'on peut effectivement se contenter de la première lecture et passer aux chapitres suivants. Il y aura très peu de références ultérieures aux théorèmes des deuxièmes lectures ; et même dans les rares cas où on le fait, on peut facilement remettre à plus tard la lecture détaillée du passage indiqué. Ceci permet de naviguer un peu dans les chapitres, même si par nature les mathématiques, à l'heure où on les découvre pour la première fois, permettent peu d'exercer cette liberté.

À savoir avant de commencer

On vous a peut-être dit, ou vous vous êtes peut-être imaginé, que les mathématiques à partir de l'université repartaient de zéro, et redéfinissaient tous les concepts utilisés. C'est à la fois vrai d'un point de vue strictement logique, et chargé d'une certaine hypocrisie.

Il est indéniable qu'une exposition préalable à des mathématiques va s'avérer indispensable à la bonne compréhension de ce qui suit. Pour prendre quelques exemples, nous ne prendrons pas le temps ici de forger l'intuition selon laquelle deux nombres réels repèrent un point dans le plan, pas plus que nous n'aurons le loisir de dessiner de nombreux vecteurs comme des flèches et de les additionner en construisant des parallélogrammes. Nous mentionnerons ces choses, mais à un rythme qui n'est raisonnable que parce que l'on vous sait déjà à l'aise avec ces concepts.

Une certaine familiarité avec le calcul est également attendue. Pensez que nous allons définir le concept de dérivée, dont vous devez avoir une connaissance informelle depuis le lycée, et que nous allons enfin *démontrer* que les fonctions ayant une dérivée positive sont croissantes ; mais apprendre à calculer les dérivées est une gymnastique à laquelle on vous suppose un peu aguerris.

Nous avons rassemblé dans un appendice, à la fin de la partie « Préliminaires », une compilation des connaissances de Terminale qui vont être les plus utiles à la lecture du présent manuel.

Remerciements

Ce livre a été principalement écrit en 2010-2011, à Vancouver, alors que l'auteur bénéficiait d'une délégation au CNRS sans laquelle il n'aurait pu venir à bout de ce projet. Depuis, de nombreux collègues ont apporté leurs remarques constructives, au premier rang desquels Vianney Combet, à qui je dois une relecture minutieuse et un nombre de corrections trop élevé pour être rendu public. Je l'en remercie vivement.

J'ai été encouragé à finaliser ce cours, qui stagnait dans une version préliminaire, par l'enthousiasme des élèves de la promotion 2011-2012 de MPA à l'université de Strasbourg. Qu'il me soit donc permis de saluer Alexandre Amara, Vincent Beck, Yann Becker, Mike Beller, Arnaud Bonnet, Béatrice Chetard, Sophie Hurier, Guillaume Klein, Kévin Pfeiffer, Philippe Ricka, Claire Roman et Étienne Werly.

Table des matières

Première partie

Préliminaires

Chapitre 1

Ensembles

Notre exploration des mathématiques commence par l'étude des ensembles, à partir desquels *tout* va être défini.

Attention, il n'est pas garanti que la meilleure chose à faire soit de commencer par ce chapitre, pour certains étudiants : si vous en ressentez le besoin, attaquez plutôt par l'appendice A, qui résume les résultats du lycée que nous supposons connus.

Les objets mathématiques peuvent être rangés dans des *ensembles*, que l'on écrit avec des accolades. Par exemple,

$$E = \{1, 2, 3\} \quad \text{et} \quad F = \{19, 11\}$$

sont des ensembles. On note $x \in X$ pour signifier que x appartient à X, et dans le cas contraire on emploie le symbole $\notin$; par exemple, on a $2 \in E$ et $3 \notin F$.

Un ensemble ne comprend jamais de « répétitions », et n'est pas ordonné : ainsi

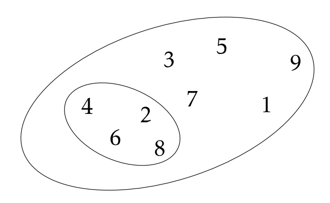
$$\{2, 2, 2, 3, 3\} = \{2, 3\} \quad \text{et} \quad \{3, 2, 1\} = \{1, 2, 3\}.$$

Il existe bien sûr des ensembles infinis, comme l'ensemble $\mathbb{N}$ des nombres entiers, dont nous reparlerons au chapitre suivant. Il y a également un ensemble vide, qui ne contient aucun élément : on le note $\emptyset$ ou, plus rarement, $\{\}$.

Lorsque tous les éléments d'un ensemble A sont aussi dans l'ensemble B, on dit que A est une *partie* de B, ou qu'il est inclus dans B, et on note $A \subset B$. Par exemple

$$\{2, 4, 6, 8\} \subset \{1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

Les ensembles sont souvent dessinés comme des bulles, et pour représenter l'inclusion on place ces bulles les unes dans les autres, comme ci-dessous :



Fixant B, on peut considérer l'ensemble $\mathcal{P}(B)$ dont les éléments sont toutes les parties de B ; ainsi dans le cas où $B = \{1, 2, 3\}$, on a

$$\mathcal{P}(B) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

(On n'oublie ni la partie vide, ni B lui-même.)

Enfin, étant donnés deux ensembles A et B, on peut former leur *produit cartésien* noté $A \times B$, dont les éléments sont les paires (a, b) avec $a \in A$ et $b \in B$. Lorsque $A = \{1, 3\}$ et $B = \{2, 4, 6\}$ par exemple, on a

$$A \times B = \{(1, 2),\ (1, 4),\ (1, 6),\ (3, 2),\ (3, 4),\ (3, 6)\}.$$

On notera que pour les paires, l'ordre est important : ainsi l'élément $(1, 2)$ de $\mathbb{N} \times \mathbb{N}$ est différent de l'élément $(2, 1)$.

On écrit volontiers A^2 au lieu de $A \times A$. Sur le même modèle on peut considérer, pour tout entier $n \geq 1$, l'ensemble noté A^n dont les éléments sont de la forme $(a_1, a_2, \ldots, a_n)$ avec chaque $a_i \in A$ (on parle des « n-uplets d'éléments de A », et pour $n = 3, 4, 5$ on parle de triplets, quadruplets, quintuplets).

voir les exercices 143, 130

Lorsqu'on dispose d'un ensemble E, on peut s'intéresser aux élements de E qui vérifient une certaine propriété P. Ceux-ci forment à nouveau un ensemble, que l'on note ainsi :

$$\{x \in \mathrm{E} \mid \mathrm{P}(x)\}.$$

(Parfois le | est remplacé par deux points, ou par l'expression complète « tels que ». Il y a de nombreuses variantes et il faut s'habituer à des notations qui changent de temps en temps, en général pour éviter les lourdeurs.)

Par exemple, supposons que $\mathrm{A} \subset \mathrm{E}$. Alors le *complémentaire de* A *dans* E est par définition

$$\{x \in \mathrm{E} \mid x \notin \mathrm{A}\}.$$

On le note généralement $\mathrm{E} - \mathrm{A}$ ou $\mathrm{E} \smallsetminus \mathrm{A}$, parfois même $\complement\mathrm{A}$ lorsque E est sous-entendu.

Autre exemple, si A et B sont deux parties de E, alors leur *intersection* est

$$\mathrm{A} \cap \mathrm{B} = \{x \in \mathrm{E} \mid x \in \mathrm{A} \text{ et } x \in \mathrm{B}\},$$

leur *union* est

$$\mathrm{A} \cup \mathrm{B} = \{x \in \mathrm{E} \mid x \in \mathrm{A} \text{ ou } x \in \mathrm{B}\}.$$

Exemple 1.1 – Prenons $\mathrm{E} = \mathbb{N} \times \mathbb{N}$, puis

$$\mathrm{A} = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid n = 0\},$$

et enfin

$$\mathrm{B} = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid m = 0\}.$$

Alors $\mathrm{A} \cap \mathrm{B} = \{(0,0)\}$. On peut également écrire

$$\mathrm{A} \cup \mathrm{B} = \{(n, m) \in \mathbb{N} \times \mathbb{N} \mid nm = 0\}.$$

Note : en pratique, on écrirait plutôt $\mathrm{A} = \{(0, m) \in \mathbb{N} \times \mathbb{N}\}$ ou encore $\mathrm{A} = \{(0, m) \mid m \in \mathbb{N}\}$, l'essentiel étant de se faire comprendre.

Il est très important de comprendre dès maintenant que la lettre x qui est employée ci-dessus dans la description des ensembles peut être remplacée par n'importe quelle autre : on obtient rigoureusement *les mêmes* ensembles. Par exemple si

$$\mathrm{A} = \{x \in \mathbb{N} \mid \text{il existe } y \in \mathbb{N} \text{ tel que } x = 2y\},$$

et si

$$\mathrm{B} = \{a \in \mathbb{N} \mid \text{il existe } b \in \mathbb{N} \text{ tel que } a = 2b\},$$

alors $\mathrm{A} = \mathrm{B} =$ les nombres entiers pairs.

voir les exercices 126, 127, 133, 135, 138, 141, 125

On ne peut pas utiliser tout et n'importe quoi pour décrire les ensembles. Pour se convaincre que les propriétés P comme ci-dessus ne peuvent pas être complètement arbitraires, voir l'encadré « Deux paradoxes ». Pour bien faire les choses, il conviendrait de définir précisément quelles sont les propriétés acceptables, ou en d'autres termes, définir ce qu'est un « énoncé mathématique ».

Deux paradoxes

L'énoncé selon lequel $\{x \in E \mid P(x)\}$ est un ensemble lorsque E est un ensemble peut paraître anodin. En réalité il est bien plus fin qu'on pourrait le croire. Nous allons voir deux paradoxes célèbres, dont l'élucidation fait intervenir de manière subtile cette construction.

Voici le premier. Pour un entier n, considérons la propriété « n ne peut pas être décrit en moins de 16 mots ». Appelons cette propriété $P(n)$, et soit

$$A = \{n \in \mathbb{N} \mid P(n)\}.$$

Les mots de la langue française sont en nombre fini, donc en 16 mots on ne peut décrire qu'un nombre fini de nombres. Ainsi, A est infini et en particulier, non-vide. Soit alors a le plus petit élément de A. Ce nombre est « le plus petit nombre qui ne peut pas être décrit en moins de 16 mots ». On vient tout juste de décrire a en 15 mots !

C'est absurde. Et pour cause, la propriété $P(n)$ ne fait pas partie des propriétés mathématiques acceptables.

Notre deuxième exemple utilise pour $P(x)$ la propriété « $x \notin x$ ». Celle-ci est parfaitement acceptable. C'est sa signification intuitive proche de zéro qui donne un parfum de paradoxe au raisonnement suivant, pourtant correct.

Montrons la chose suivante : pour tout ensemble E, il existe un ensemble A tel que $A \notin E$. En effet, soit

$$A = \{x \in E \mid x \notin x\}.$$

Si on avait $A \in E$, alors on constaterait que $A \in A$ équivaut à $A \notin A$, par définition. C'est absurde, donc $A \notin E$.

On énonce souvent ce résultat sous la forme suivante : il n'existe pas d'ensemble de tous les ensembles. Nous venons bien de le démontrer. S'il est tentant d'écrire quelque chose comme $U = \{x \mid x \text{ est un ensemble}\}$ pour essayer de le définir malgré tout, on se rend compte que cette expression n'est pas de la forme $\{x \in E \mid P(x)\}$, et donc ne désigne pas un ensemble. La présence de l'ensemble E pour « chapeauter » les x est essentielle.

Cette théorie existe, et il existe même plusieurs systèmes concurrents. Cependant il serait complètement hors de propos de donner une description précise de l'un de ces systèmes dès maintenant (les détails sont parfois donnés en troisième ou quatrième année, et encore). Nous allons nous contenter d'une discussion informelle qui suit les grandes lignes de ce que l'on appelle *la logique du premier ordre* (pour des raisons que l'on n'expliquera pas).

Nous avons rencontré des propositions mathématiques : $x \in A$ par exemple, et on pourrait citer aussi les égalités comme $x = y$. La négation d'une proposition en est une, ainsi $x \notin A$ est un énoncé mathématique.

On peut créer de nouveaux énoncés à l'aide de « ou » et de « et » : nous l'avons fait dans la définition des intersections et des unions. On peut aussi relier deux énoncés P et Q par le symbole $\Rightarrow$, qui se lit « implique ». On obtient l'énoncé $P \Rightarrow Q$, qui est faux lorsque P est vrai et Q est faux ; dans tous les autres cas $P \Rightarrow Q$ est vrai. Voyons un exemple :

$$A = \{(x, y) \in \mathbb{N} \times \mathbb{N} \mid x \neq 0 \Rightarrow y = 0\}.$$

Les éléments de A sont les paires $(x, 0)$ avec x entier, ainsi que les paires $(0, y)$ avec y entier.

Le symbole $\Rightarrow$ est surtout pertinent lorsqu'on l'utilise en conjonction avec le *quantificateur universel*, c'est-à-dire le petit symbole $\forall$ qui signifie « pour tout ». Nous pouvons par exemple utiliser ce symbole pour montrer que $A \subset B$ est un énoncé mathématique : en effet il revient à dire

$$\forall x,\ x \in A \Rightarrow x \in B.$$

L'autre quantificateur à notre disposition est le *quantificateur existentiel*, qui s'écrit $\exists$ et signifie « il existe ». On a déjà observé que, pour un nombre entier n, la propriété « n est pair » s'écrit

$$\exists m \in \mathbb{N} \text{ tel que } n = 2m.$$

(En toute rigueur, en logique du premier ordre on écrit plutôt $\exists m,\ m \in \mathbb{N}$ et $n = 2m$. On s'autorise un peu de souplesse pour plus de clarté.)

En règle générale, un « énoncé mathématique » est une phrase que l'on peut réduire à une suite de symboles combinant $\forall$, $\exists$, $\in$, $=$, $\Rightarrow$, des négations, des « ou » et des « et ». En pratique cependant, la moindre définition, le moindre théorème, occuperaient des milliers de symboles si on voulait les décortiquer complètement. En conséquence, il faut veiller en permanence à ce que les énoncés que l'on produit soient *théoriquement* remplaçables par des symboles, sans jamais effectuer concrètement ce remplacement. Notons tout de même qu'à l'aide d'un ordinateur, on peut parfois rédiger certaines démonstrations jusqu'au moindre détail : c'est ce qu'on appelle les « preuves certifiées ».

Ajoutons enfin que dans certaines situations, nous utiliserons les symboles $\forall$, $\exists$ ou autres, lorsque l'on souhaite lever toute ambiguïté. Ainsi de la définition des limites, par exemple. *voir les exercices 110, 112*

Étant donnés deux ensembles A et B, une fonction f de A vers B associe à tout élément $x \in A$ un élément $f(x) \in B$ et un seul. On peut traduire cette définition (un peu vague) en termes d'ensembles. Si l'on souhaite être extrêmement précis, on dira :

DÉFINITION 1.2 – Une *fonction*, ou *application*, est un objet f déterminé par trois ensembles :

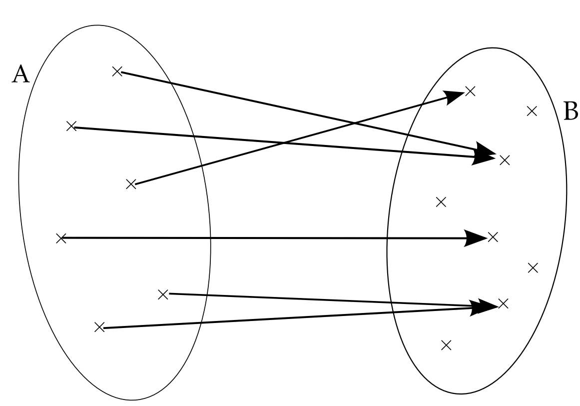
1. un ensemble A, appelé le domaine de définition de f, ou parfois la source de f ;
2. un ensemble B, appelé le but de f ;
3. un ensemble Γ, qui est une partie de $A \times B$ et que l'on appelle *le graphe de* f, ayant la propriété suivante : pour chaque $x \in A$, il existe un unique $y \in B$ tel que $(x, y) \in \Gamma$. Ce y est noté $f(x)$.

On utilise la notation

$$f : A \longrightarrow B$$

pour indiquer que f est une fonction dont le domaine de définition est A et dont le but est B.

On représente typiquement une fonction $A \to B$ de la manière suivante :



Chaque flèche sur ce dessin part d'un élément $x \in A$ et pointe sur $f(x)$. La caractéristique importante est que chaque point de A marque le début d'une flèche, et d'une seule.

Voyons quelques exemples.

EXEMPLE 1.3 – Il y a une (et une seule) fonction $f : \mathbb{N} \to \mathbb{N}$ telle que $f(n) = 2n^2 + 1$. On utilise parfois la notation

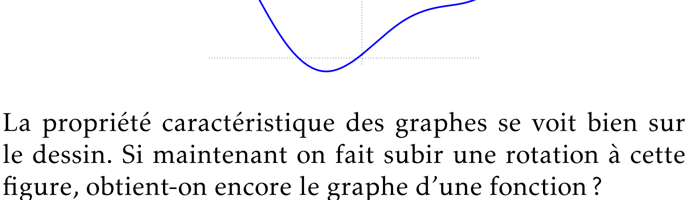
$$\begin{array}{rcl} f : \mathbb{N} & \longrightarrow & \mathbb{N} \\ n & \mapsto & 2n^2 + 1 \end{array}$$

pour désigner cette fonction. C'est très souvent par des formules, telles que $2n^2 + 1$, que l'on va définir les fonctions.

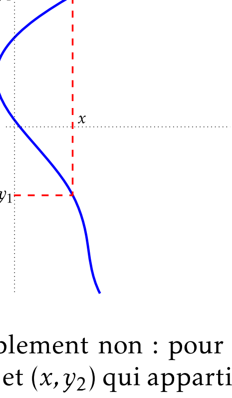
Ici le domaine de définition est $A = \mathbb{N}$, le but est $B = \mathbb{N}$, et le graphe de f est $\Gamma = \{(n, 2n^2 + 1) \mid n \in \mathbb{N}\}$.

EXEMPLE 1.4 – Soit $p \colon \mathbb{N} \smallsetminus \{0\} \to \mathbb{N}$ la fonction telle que $p(n) =$ le n-ième nombre premier. Ainsi $p(1) = 2$, $p(2) = 3$, $p(3) = 5$, $p(4) = 7$ et ainsi de suite. Cette fonction p est bien définie, même si on n'a pas utilisé de formule. (Cela dit, il en existe.)

EXEMPLE 1.5 – Nous allons anticiper un peu et supposer que vous connaissez un minimum l'ensemble $\mathbb{R}$. On le représente par une droite, et $\mathbb{R} \times \mathbb{R}$ par un plan. Une fonction $A \to B$ avec $A \subset \mathbb{R}$ et $B \subset \mathbb{R}$ est donnée par son graphe, qui ressemble de près ou de loin à une courbe dans le plan. Par exemple la figure suivante représente un tel graphe.



La propriété caractéristique des graphes se voit bien sur le dessin. Si maintenant on fait subir une rotation à cette figure, obtient-on encore le graphe d'une fonction ?



La réponse est visiblement non : pour le x indiqué, il y a deux couples (x, y_1) et (x, y_2) qui appartiennent à la courbe. Ce n'est donc pas un graphe. On retiendra la traduction géométrique simple : lorsque $A \subset \mathbb{R}$ et $B \subset \mathbb{R}$, une partie Γ de $A \times B$ est le graphe d'une fonction $A \to B$ si et seulement si chaque droite verticale d'équation $x = a$ (avec $a \in A$) coupe Γ exactement en un point.

Dans la suite du chapitre nous allons étudier la propriété correspondante en utilisant cette fois des droites horizontales. *voir l'exercice 132*

– Fin de la première lecture –

Définition 1.6 – Soit $f : A \to B$ une fonction. Supposons que, pour tout choix de deux éléments distincts $x_1 \neq x_2$ dans l'ensemble A, on ait également $f(x_1) \neq f(x_2)$. On dit alors que f est *injective*, ou encore que f est une *injection*.

Il existe bien des façons de reformuler ceci. Par exemple, f est injective si et seulement si l'égalité $f(x_1) = f(x_2)$ entraîne $x_1 = x_2$. Également, il est bon de noter que f est injective si et seulement si l'équation

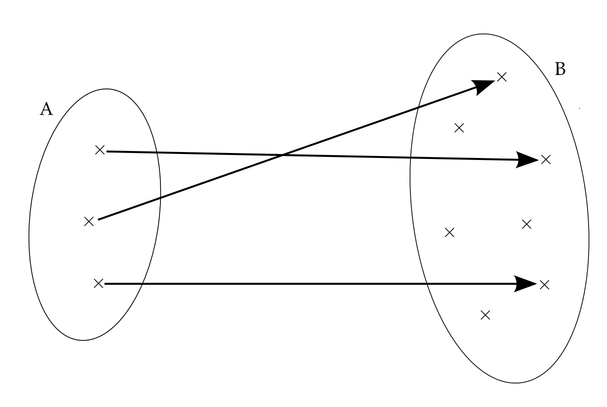
$$f(x) = b\,,$$

dont l'inconnue est $x \in A$ et qui comporte le paramètre $b \in B$, possède *au maximum* une solution.

Exemple 1.7 – La fonction $d : \mathbb{N} \to \mathbb{N}$ définie par $d(n) = 2n$, est injective : en effet si $2x_1 = 2x_2$, alors $x_1 = x_2$. L'équation $d(x) = b$ s'écrit $2x = b$; elle a une solution $x = \frac{b}{2}$ si b est pair, et aucune solution si b est impair.

Exemple 1.8 – La fonction $c : \mathbb{Z} \to \mathbb{N}$ définie par $c(n) = n^2$ n'est *pas* injective (ici $\mathbb{Z}$ est l'ensemble de tous les nombres entiers, positifs ou négatifs). En effet $c(n) = c(-n)$, de sorte que l'équation $c(x) = b$, qui s'écrit $x^2 = b$, peut posséder deux solutions, comme par exemple 2 et -2 qui sont solutions pour $b = 4$.

Voici comment on représente une fonction injective :

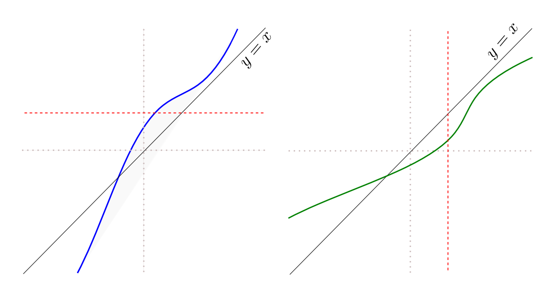


Cette fois-ci, les flèches pointent toutes vers des éléments différents.

Exemple 1.9 – Revenons au cas particulier où $f : A \to B$ avec A et B des parties de $\mathbb{R}$. L'équation $f(x) = b$ possède une solution x lorsque le graphe de f comporte un point $(x, f(x))$ qui est également sur la droite horizontale d'équation $y = b$. La condition pour que f soit injective est donc que les droites horizontales rencontrent le graphe de f en un point au maximum.

Soit Γ le graphe de f. Faisons subir à ce graphe une symétrie par rapport à la droite d'équation $y = x$ (cette symétrie envoie le point (x, y) sur (y, x)). On obtient un ensemble Γ'. Lorsque f est injective, ce Γ' ne rencontre les droites verticales qu'en un point au plus. C'est-à-dire que Γ' est le graphe d'une fonction !

Cette discussion est illustrée sur la figure suivante.



Soyons plus précis. Pour définir une fonction g dont le graphe serait Γ', il lui faut un ensemble de définition et un but. Les points de Γ' sont ceux de la forme $(f(x), x)$. Notons donc

$$f(A) = \{f(x) \mid x \in A\} \subset B.$$

(Nous reviendrons sur cette notation (abrégée) dans le paragraphe suivant.) Alors on peut définir une fonction $g : f(A) \to A$ dont le graphe est Γ'. Concrètement, on a $g(f(a)) = a$, ce qui a un sens puisque f est injective.

Cette fonction g est essentiellement ce qu'on appelle la *réciproque de* f, qui se note f^{-1}. Toutefois il nous reste un peu de vocabulaire à introduire avant de détailler ceci.

Définition 1.10 – Soit $f : A \to B$ une fonction. On note $f(A)$, ou encore $\mathrm{Im}(f)$, l'ensemble

$$\{b \in B \mid \exists x \in A \text{ tel que } b = f(x)\}.$$

(En plus concis $f(A) = \{f(x) \mid x \in A\}$.) On dit que $f(A)$ est l'image de A par f.

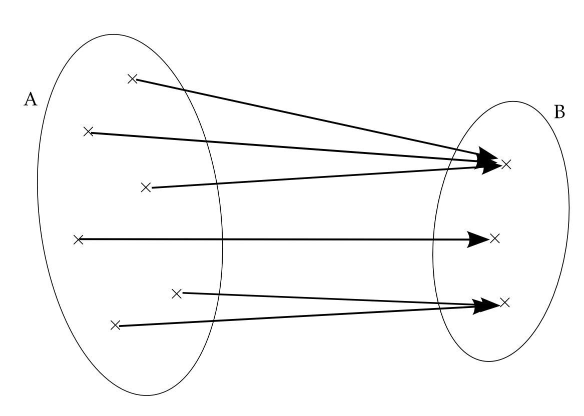
Lorsque $f(A) = B$, on dit que f est *surjective*, ou encore que f est une *surjection*.

Ainsi f est surjective lorsque l'équation $f(x) = b$ possède *au minimum* une solution.

Exemple 1.11 – La fonction $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ définie par $f(n, m) = n + m$ est surjective. En effet, si on se donne $b \in \mathbb{N}$, alors $f(b, 0) = b$. On a aussi $f(0, b) = b$, et même $f(1, b - 1) = b$, de sorte que f est loin d'être injective, par contre.

Exemple 1.12 – La fonction $d : \mathbb{N} \to \mathbb{N}$ telle que $d(n) = 2n$ n'est pas surjective. En fait l'ensemble image $d(\mathbb{N})$ est l'ensemble des nombres pairs.

Voici la représentation typique d'une fonction surjective :



Ici chaque élément de B est à l'extrémité d'au moins une flèche.

Définition 1.13 – Lorsqu'une fonction est à la fois injective et surjective, on dit qu'elle est *bijective*, ou encore que c'est une *bijection*.

Lorsque $f : A \to B$ est bijective, l'équation $f(x) = b$ possède *une solution et une seule* (comme précédemment, l'inconnue est x et b est un paramètre). Cette solution est notée $f^{-1}(b)$.

On obtient ainsi une fonction $f^{-1} : B \to A$, que l'on appelle la *réciproque* de f. On a alors :

Proposition 1.14 – *Lorsque f est bijective, la fonction f^{-1} vérifie*

1. $f^{-1}(f(a)) = a$ *pour* $a \in A$,
2. $f(f^{-1}(b)) = b$ *pour* $b \in B$.

Réciproquement si on a une paire de fonctions $f : A \to B$ et $g : B \to A$ telles que $g(f(a)) = a$ pour $a \in A$ et $f(g(b)) = b$ pour $b \in B$, alors f est une bijection et $g = f^{-1}$.

Enfin, f^{-1} est également une bijection lorsqu'elle existe, et

$$(f^{-1})^{-1} = f.$$

Démonstration. (1) Étant donné a, soit $b = f(a)$; puisque f est injective a est le seul élément de A qui vérifie cette équation, et c'est cet élément que l'on note $f^{-1}(b)$. Donc $a = f^{-1}(b) = f^{-1}(f(a))$.

(2) C'est la définition même de $f^{-1}(b)$.

Montrons la réciproque. Soient f et g comme dans la proposition. Si $f(a_1) = f(a_2)$, alors on a aussi $g(f(a_1)) = g(f(a_2))$, donc $a_1 = a_2$. Ainsi f est injective. De plus, si $b \in B$ on a $b = f(g(b))$ donc b est bien dans l'image de f, ce qui montre que f est surjective. Finalement f est une bijection.

Partant de $f(f^{-1}(b)) = b = f(g(b))$, on applique g pour obtenir

$$g[f(f^{-1}(b))] = g[f(g(b))].$$

Puisque $g(f(a)) = a$ pour tout $a \in A$ (et donc en particulier pour $a = f^{-1}(b)$ ou pour $a = g(b)$), cette dernière égalité se simplifie et donne $f^{-1}(b) = g(b)$. Donc $f^{-1} = g$.

Par symétrie, on peut inverser les rôles de f et de g. Donc g est bijective et $g^{-1} = f$, c'est-à-dire que f^{-1} est bijective et que $(f^{-1})^{-1} = f$. □

Exemple 1.15 – La fonction $s : \mathbb{Z} \to \mathbb{Z}$ définie par $s(n) = -n$ est une bijection. De plus, $s^{-1} = s$.

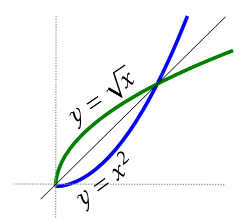
voir les exercices 187, 190, 201

Vous connaissez depuis le lycée quelques paires de bijections réciproques : exponentielle et logarithme, cosinus et arccosinus, sinus et arcsinus, tangente et arctangente. Des rappels complets sont données dans l'appendice A, voir la partie « Fonctions usuelles » en page 140.

Attardons-nous ici sur l'exemple très simple des fonctions « carré » et « racine carrée ». Plus précisément, la fonction

$$f : [0,+\infty[\to [0,+\infty[$$

définie par $f(x) = x^2$ est une bijection. Sa réciproque f^{-1} s'appelle la fonction « racine carrée » et se note $f^{-1}(x) = \sqrt{x}$.



Que sommes-nous capables de véritablement démontrer ? Commençons par l'injectivité de f. Si $f(x_1) = f(x_2)$, on a $x_1^2 = x_2^2$, d'où

$$x_1^2 - x_2^2 = (x_1 - x_2)(x_1 + x_2) = 0\,.$$

Or puisqu'on se restreint à $x_1 \geq 0$ et $x_2 \geq 0$, on ne peut avoir $x_1 + x_2 = 0$ que lorsque $x_1 = x_2 = 0$. Dans les autres cas, on simplifie par $x_1 + x_2$ et on en conclut que $x_1 = x_2$, là encore. Donc f est injective.

La fonction f, dont le but est $[0,+\infty[$ est-elle bien surjective ? C'est une question bien plus difficile ! Il s'agit de savoir si tout nombre réel $b \geq 0$ possède « une racine carrée », c'est-à-dire s'il existe x tel que $b = x^2$. En d'autres termes, est-ce qu'on peut toujours donner un sens à la notation $\sqrt{b}$? Bien sûr nous venons d'affirmer ci-dessus que la réponse est oui, mais comment le démontrer ?

C'est l'objet du chapitre suivant, et c'est aussi notre première rencontre avec un énoncé considéré comme évident jusqu'au lycée et qu'il va falloir élucider. L'étude des « fonctions usuelles » en contient bien d'autres (qu'est-ce que l'exponentielle, au juste ? qu'est-ce qu'un cosinus ? etc.).

DÉFINITION 1.16 – Soit $f\colon E \to F$ et $g\colon F \to G$ deux fonctions. La *composée* de f et g, notée $g \circ f$, est la fonction $E \to G$ définie par $g \circ f(x) = g(f(x))$ pour $x \in E$.

EXEMPLE 1.17 – Prenons $E = F = G = \mathbb{R}$, puis $f(x) = x + 1$ et $g(x) = x^2$. Alors $g \circ f(x) = x^2 + 2x + 1$.

Pour finir, nous allons maintenant donner un sens à $f(A)$ lorsque A est une partie du domaine de définition de f, ainsi qu'à $f^{-1}(B)$ lorsque B est une partie du but de f. *Attention*, la notation $f^{-1}(B)$ va prendre un sens même lorsque f n'est pas une bijection (et que la notation f^{-1} toute seule n'a pas de sens). Plus précisément :

DÉFINITION 1.18 – Soit $f\colon E \to F$ une application. Pour $A \subset E$ on note

$$f(A) = \{y \in F \mid y = f(x) \text{ pour un } x \in A\} \subset F.$$

Pour $B \subset F$, on définit

$$f^{-1}(B) = \{x \in A \mid f(x) \in B\} \subset E.$$

EXEMPLE 1.19 – Pour $c\colon \mathbb{Z} \to \mathbb{N}$ telle que $c(n) = n^2$, on constate que $c(\{1,2,3\}) = \{1,4,9\}$, et $c^{-1}(\{1,4,9\}) = \{-3,-2,-1,1,2,3\}$. (La fonction c^{-1} n'existe pas.)

EXEMPLE 1.20 – Les ensembles de la forme $f^{-1}(\{y\})$ pour $y \in F$ sont souvent appelés les *fibres* de f. Avec ce langage on énonce que f est surjective si et seulement si ses fibres sont toutes non-vides, et f est injective si et seulement si ses fibres ont au plus un élément.

Lorsque f est bijective, et que la fonction f^{-1} existe, on a $f^{-1}(\{y\}) = \{f^{-1}(y)\}$.

voir les exercices 185, 186, 193, 194, 199, 124

Nous arrivons au terme de ce premier chapitre. Il reste une question difficile : avons-nous vraiment commencé par le commencement ?

S'il existe une distinction essentielle entre les mathématiques (en tout cas dans la vision idéalisée qu'on peut en avoir) et la plupart des autres disciplines, c'est sans doute qu'on y a tout le loisir de poser des questions. Qu'on essaie de demander à un physicien la définition d'une force, ou la définition de l'énergie (et non pas la formule qui calcule telle ou telle incarnation de l'énergie), et on rencontrera rapidement des difficultés, qui sont profondes et inévitables. Richard Feynman dans son « Cours de Physique » donne une belle définition de l'énergie, par ailleurs très mathématique et sans doute décevante par certains égards pour les physiciens. Il ne parvient pas à en faire autant pour les forces, et il est intéressant de lire ses explications.

Richard Feynman, Le cours de Physique de Feynman, *Dunod, 1999.*

En théorie, ceci n'arrive jamais en mathématiques. Vous pouvez demander à votre professeur de définir ce qu'est le logarithme, il le fera (par exemple) en disant que c'est une intégrale ; vous pouvez demander ce qu'est une intégrale, vous aurez une réponse qui fait intervenir des limites ; vous pouvez ensuite demander ce que signifie un « passage à la limite », etc. Mais que va-t-il arriver lorsqu'on en finit par demander ce qu'est un ensemble, ce que sont les nombres entiers, et pourquoi $2+2=4$? Il va bien falloir trouver une réponse.

Cependant, a-t-on vraiment le désir de traiter cette question *maintenant*, dans le premier chapitre d'un livre destiné aux étudiants en première année ? Nous affrontons un véritable dilemme. D'un côté, par simple honnêteté (et pas seulement pour avoir des réponses à disposition d'un étudiant qui aurait l'idée de demander la définition des choses « évidentes »), on a bien envie de commencer par le commencement, et de définir tous les objets que l'on rencontre en partant « de rien ». D'un autre côté, on peut objecter que cette exigence serait aussi déraisonnable que d'imposer à chaque candidat au permis de conduire de connaître entièrement la mécanique automobile avant même sa première heure de conduite.

De fait, la vaste majorité des mathématiciens de profession ne connaissent pas et ne souhaitent pas connaître les détails des fondements logiques des mathématiques. Ils en connaissent cependant les grandes lignes, que nous allons exposer tout de suite.

Le principe de départ de la « méthode axiomatique » est simple. On *postule* l'existence de certains objets, vérifiant certaines propriétés appelées axiomes. Par « postuler », il faut comprendre qu'il s'agit de se donner des règles du jeu, que l'on accepte sans les questionner. Ensuite, les résultats que l'on peut démontrer à partir de ces axiomes sont considérés comme « vrais dans la théorie ».

Le premier exemple remonte à l'Antiquité, c'est celui des axiomes d'Euclide pour la géométrie. Euclide postule l'existence d'objets appelés points et droites (et d'autres encore), sachant qu'un point peut « appartenir » à une droite. Ceci dans le respect de certaines propriétés, comme « deux droites parallèles à une même troisième sont parallèles » (et bien sûr, dans cette théorie l'expression « être parallèles » est elle-même définie, à l'aide de concepts premiers comme l'appartenance d'un point à une droite). Toute la géométrie est déduite de ces axiomes.

En principe, comme le disait Hilbert, on pourrait remplacer « point » par « table », « droite » par « chaise », et « appartenir » par n'importe quel verbe, et on pourrait toujours développer la théorie, de manière purement formelle. Ceci est vrai ; ce ne sont que des mots. Toutefois, il faut se garder de prendre ceci trop au sérieux : les axiomes ont été choisis parce qu'Euclide a l'intuition que le monde réel comporte des points et des droites (ou au moins des segments), et parce qu'il souhaite considérer chaque résultat « vrai dans la théorie » comme une assertion vraie sur le monde réel.

L'avantage de la méthode axiomatique est de couper court aux débats sur l'existence des objets de départ. On suppose qu'ils existent, vérifiant certaines propriétés, le reste n'est que déduction. Celui qui doute de l'existence de ces objets peut entrer dans un débat philosophique, par ailleurs intéressant, mais il ne peut pas critiquer le travail mathématique de ceux qui ont choisi ces axiomes (sauf à montrer que les axiomes sont contradictoires et que l'on peut en déduire des choses absurdes, comme un énoncé et son contraire simultanément, par exemple).

On continue de nos jours à employer la méthode axiomatique, même si les mathématiques modernes ne reposent plus sur les axiomes d'Euclide. Il existe plusieurs systèmes d'axiomes possibles, et dans l'optique de ce livre il n'est absolument pas utile d'en comprendre les différences, ni même d'en décrire un en détail. Citons tout de même :

1. Le système de *l'arithmétique de Peano*. On choisit ici de prendre les nombres entiers comme objets de départ, et on suppose qu'ils vérifient certaines propriétés comme « tout nombre n possède un successeur $n+1$ ». On déduit tout le reste.

2. La *théorie des ensembles de Zermelo & Fraenkel*. Les objets de départ sont les ensembles et les axiomes sont, en gros, les propriétés décrites dans la première partie de ce chapitre.

3. Il existe aussi un système qui part des *fonctions* comme objets primaires.

Les théorèmes que l'on peut obtenir dans un système sont en général démontrables dans les autres systèmes. Ce n'est pas *toujours* vrai, et on obtient des résultats un peu plus forts avec la théorie des ensembles qu'avec l'arithmétique ; mais les différences sont subtiles et nous n'en parlerons pas plus. Ceci signifie qu'étant donné un système de départ, disons la théorie des ensembles, il faut pouvoir définir les objets des autres systèmes, comme les nombres entiers ou les fonctions.

Dans ce cours, on ne va pas s'encombrer de telles considérations, et nous considérerons comme « connus » aussi bien les ensembles que les nombres entiers.

Chapitre 2

Nombres

Dans ce chapitre nous allons faire connaissance avec les systèmes de nombres $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ et $\mathbb{C}$. Le plat de résistance est $\mathbb{R}$, qui est d'ailleurs trop compliqué pour qu'on vous dise tout à son sujet ! Cependant nous allons mettre en valeur sa propriété essentielle, qui est celle de la « borne supérieure ».

Dans la deuxième partie, qui relève plus de l'algèbre, nous décrivons quelques « nombres » plus originaux.

Ce chapitre est destiné aux étudiants en filière mathématique, et va moins intéresser les élèves chimistes ou ingénieurs, pour qui l'intuition que l'on a de $\mathbb{R}$ depuis le lycée devrait suffire. Si vous ne lisez pas ce chapitre, il y aura simplement dans la suite quelques démonstrations que vous ne pourrez pas comprendre (en particulier celle de l'énoncé sur les suites croissantes et majorées dans le chapitre « Suites », ou la construction de l'intégrale de Riemann dans le chapitre du même nom).

Le premier ensemble de nombres à notre disposition est celui des nombres naturels :

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

Puis vient l'ensemble des nombres relatifs $\mathbb{Z}$, qui contient $\mathbb{N}$, et comprend également les nombres négatifs comme -4. Enfin nous avons l'ensemble des nombres rationnels $\mathbb{Q}$, c'est-à-dire l'ensemble des fractions $\frac{p}{q}$ avec $p, q \in \mathbb{Z}$ et $q \neq 0$. Noter les inclusions $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$.

Dans le chapitre précédent nous avons expliqué que nous ne définirons pas l'ensemble $\mathbb{N}$, considéré comme naturel (d'où son nom). Par contre on peut parfaitement donner une définition des ensembles $\mathbb{Z}$ et $\mathbb{Q}$ à partir de $\mathbb{N}$: voir l'encadré « Une définition de $\mathbb{Q}$ ». Quoi qu'il en soit, nous pouvons considérer que nous sommes à l'aise avec les nombres rationnels.

Une définition de $\mathbb{Q}$

Imaginons quelqu'un qui connaisse l'ensemble $\mathbb{Z}$ mais pas $\mathbb{Q}$: comment le lui décrire ? (À titre d'exercice vous pourrez ensuite décrire $\mathbb{Z}$ à quelqu'un qui connaît $\mathbb{N}$).

On peut facilement imaginer *définir* une fraction comme étant une paire de nombres $(p, q) \in \mathbb{Z} \times \mathbb{Z}$ avec $q \neq 0$, avec la convention que (p, q) et (a, b) représentent la même fraction lorsque $bp = aq$, puisque

$$\frac{p}{q} = \frac{a}{b} \Leftrightarrow bp = aq.$$

En étant tout à fait précis, on est amené à la définition suivante, étonnamment compliquée : étant donnée une paire (p, q) de nombres avec $q \neq 0$, la *fraction* définie par ce couple est l'ensemble

$$F_{p,q} = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \neq 0 \text{ et } pb = aq\}.$$

On décide d'écrire $\frac{p}{q}$ au lieu de $F_{p,q}$, par simplicité.

Maintenant si (a, b) vérifie $pb = aq$, on peut *démontrer* que

$$\frac{p}{q} = \frac{a}{b}.$$

Faisons-le : montrons que $F_{p,q} = F_{a,b}$. C'est une égalité d'ensembles ! Soit donc (x, y) un couple de nombres entiers avec $y \neq 0$. Si $(x, y) \in F_{p,q}$, on a $py = xq$. Multipliant par b, on obtient $pby = xbq$. Or on a supposé que $pb = aq$, donc on a $aqy = xbq$. En simplifiant par q qui est non-nul, on en tire $ay = xb$, c'est-à-dire $(x, y) \in F_{a,b}$. Ceci montre que $F_{p,q} \subset F_{a,b}$; cet argument est visiblement symétrique, donc de la même manière on a $F_{a,b} \subset F_{p,q}$, et on conclut que $F_{p,q} = F_{a,b}$ comme on le souhaitait.

Réciproquement, comme $(a, b) \in F_{a,b}$, l'égalité $F_{p,q} = F_{a,b}$ entraîne $pb = aq$.

Nous avons donc donné une définition du symbole $\frac{p}{q}$ qui obéit à au moins une règle que nous attendons, la règle du « produit en croix ».

Pour définir $\mathbb{Q}$, il reste du travail : il faut expliquer l'addition et la multiplication.

On pourrait croire que c'est facile. Soient F_1 et F_2 deux fractions. Choisissons (p, q) tels $F_1 = \frac{p}{q}$ (c'est possible par définition d'une fraction), puis choisissons (a, b) tels que $F_2 = \frac{a}{b}$.

On pose alors

$$F_1 + F_2 = \frac{pb + aq}{qb},$$

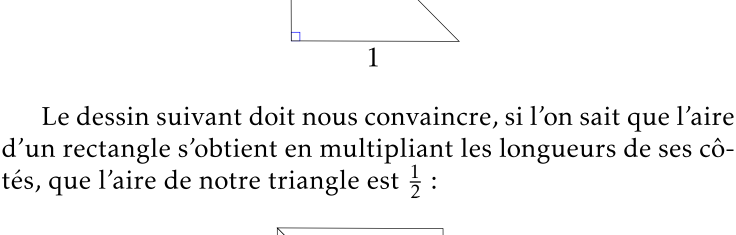
et

$$F_1 \times F_2 = \frac{pa}{qb}.$$

(On fait ceci évidemment en pensant aux formules pour $\frac{p}{q} + \frac{a}{b}$ et $\frac{p}{q} \times \frac{a}{b}$.)

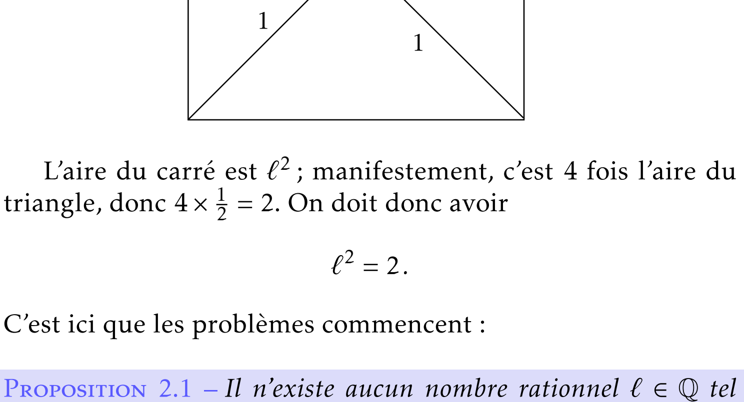
Malheureusement il reste des vérifications à faire : il faut bien s'assurer que le résultat ne dépend pas des choix que nous avons été obligés de faire pour (p, q) et (a, b), qui ne sont pas les seuls représentants de leur fraction. Nous laissons ces détails au lecteur.

A-t-on besoin d'autres nombres que des rationnels ? C'est une question qui remonte aux Grecs de l'Antiquité. Les difficultés apparaissent à peu près ainsi. Les nombres doivent au minimum être capables de mesurer les aires et les longueurs des objets qui nous entourent (c'est un petit anachronisme car les Grecs ne pensaient pas (encore) aux aires comme à des nombres, mais l'idée est la même). Imaginons donc un triangle rectangle et isocèle, dont le petit côté est de longueur 1, comme ci-dessous.



Le dessin suivant doit nous convaincre, si l'on sait que l'aire d'un rectangle s'obtient en multipliant les longueurs de ses côtés, que l'aire de notre triangle est $\frac{1}{2}$:

Maintenant, notons ℓ la longueur de l'hypoténuse (le grand côté du triangle), et considérons ce dernier dessin, obtenu à partir de 4 copies du triangle initial :



L'aire du carré est ℓ^2 ; manifestement, c'est 4 fois l'aire du triangle, donc $4 \times \frac{1}{2} = 2$. On doit donc avoir

$$\ell^2 = 2.$$

C'est ici que les problèmes commencent :

Proposition 2.1 – *Il n'existe aucun nombre rationnel $\ell \in \mathbb{Q}$ tel que $\ell^2 = 2$.*

Démonstration. Supposons par l'absurde que l'on ait $\ell = \frac{p}{q}$ tel que $\ell^2 = 2$, donc $p^2 = 2q^2$. Quitte à simplifier la fraction un certain nombre de fois par 2, on peut supposer que p et q *ne sont pas tous les deux pairs*.

Maintenant si l'on observe la relation $p^2 = 2q^2$, on voit que p^2 est pair ; donc p est pair également, ce que l'on va écrire $p = 2r$. Par suite $p^2 = 4r^2 = 2q^2$, donc $q^2 = 2r^2$.

On en conclut que q^2 est pair, donc q aussi. C'est une contradiction. □

Que faut-il en conclure ? Tout simplement que les nombres rationnels ne sont pas assez compétents pour décrire le monde réel. Pour être plus précis, si l'on veut assigner des nombres aux longueurs et aux aires, de sorte que certaines propriétés souhaitables soient satisfaites (par exemple en s'assurant que l'aire d'un rectangle est le produit des longueurs), alors on ne peut pas utiliser (seulement) les nombres rationnels.

Nous allons donner un survol de la construction de l'ensemble $\mathbb{R}$ des nombres réels. La discussion va rester un peu informelle et nous allons omettre un certain nombre de détails. Plus loin dans ce chapitre nous indiquerons de manière extrêmement précise comment on peut travailler avec $\mathbb{R}$ en ignorant ces détails.

Commençons par rappeler l'écriture décimale, que vous connaissez depuis longtemps : il s'agit d'écrire par exemple $572,413$ pour

$$572 + \frac{4}{10} + \frac{1}{100} + \frac{3}{1000}.$$

Par ailleurs, bien sûr, lorsqu'on écrit 572 c'est aussi un raccourci pour $5\times 100 + 7\times 10 + 2$, et l'écriture décimale consiste simplement à autoriser les puissances négatives de 10, qui démarrent à partir de la virgule.

Les nombres que l'on peut écrire ainsi sont appelés les *nombres décimaux* ; notez qu'on dit parfois un « décimal » au lieu d'un « nombre décimal », et que d'autre part un nombre décimal comporte des *chiffres décimaux* (dans l'exemple ce sont 5 ; 7 ; 2 ; 4 ; 1 ; 3) ; enfin les chiffres décimaux après la virgule sont parfois appelés les « décimales » (au féminin) du nombre. Attention à tout ce vocabulaire !

Les nombres décimaux sont des éléments de $\mathbb{Q}$, puisque, pour reprendre l'exemple, on a $572,413 = \frac{572413}{1000}$. Par contre tous les éléments de $\mathbb{Q}$ ne sont pas des décimaux. Par exemple, montrons qu'on ne peut pas écrire $\frac{1}{3}$ sous forme décimale : en effet si c'était le cas, on pourrait écrire $\frac{1}{3} = \frac{a}{10^k}$ avec $a \in \mathbb{Z}$, d'où $10^k = 3a$. Or c'est absurde puisque 3 ne divise pas $10^k = 2^k 5^k$ (nous reviendrons sur les questions de divisibilité plus loin dans ce livre ; si vous connaissez la règle qui affirme que les multiples de 3 ont la propriété que la somme de leurs chiffres est également un multiple de 3, alors vous voyez bien que ça ne marche pas pour $10^k = 100\ldots 00$ puisque la somme de ses chiffres est 1). Donc $\frac{1}{3}$ n'est pas un nombre décimal.

Voici une définition temporaire : l'ensemble $\mathbb{R}$ est l'ensemble des écritures décimales avec une *infinité* de chiffres décimaux après la virgule. Donc un élément de $\mathbb{R}$ par définition est une écriture (symbolique)

$$n, a_1 a_2 a_3 \ldots$$

où $n \in \mathbb{Z}$ et $a_i \in \{0,1,2,3,4,5,6,7,8,9\}$ pour $i \geq 1$.

Une difficulté immédiate apparaît. Imaginons savoir faire avec les éléments de $\mathbb{R}$ toutes les opérations usuelles d'addition, de multiplication, etc, de sorte que les règles de calcul auxquelles nous sommes habitués soient valides (on vous laisse déterminer quelles règles on utilise dans l'argument ci-dessous). Nous allons faire alors une découverte étonnante en considérant le nombre

$$x = 0,9999999\ldots \in \mathbb{R}.$$

En effet on doit avoir

$$10x = 9,999999\ldots$$

et donc

$$9x = 10x - x = 9,0000\ldots = 9.$$

Finalement en divisant par 9 nous avons $x = 1$. Donc

$$0,99999\ldots = 1.$$

Deux écritures différentes peuvent donc représenter le même élément (et il va falloir changer la définition de $\mathbb{R}$ puisque $0,999\ldots$ n'est pas la même écriture symbolique que $1,000\ldots$). À y bien réfléchir, ça n'est pas plus surprenant que l'égalité

$$\frac{1}{2} = \frac{5}{10};$$

nous avons déjà rencontré des écritures distinctes pour le même nombre. (Par ailleurs, si vous n'arrivez pas à convaincre quelqu'un que le nombre x est égal à 1, dites-lui que si on avait $x < 1$ alors on pourrait trouver y tel que $x < y < 1$, par exemple la moyenne $y = \frac{x+1}{2}$; on se demande bien quelles seraient alors les décimales de y !)

Évidemment on obtient de la même manière des égalités similaires, comme

$$1,25 = 1,2499999\ldots$$

par exemple.

Il serait naturel à ce stade de penser que bien d'autres bizarreries vont se produire avec cette définition de $\mathbb{R}$, hormis le « problème des infinités de 9 ». On imagine qu'en cherchant un peu on va trouver tout un tas d'égalités entre des écritures très différentes ! Pourtant, il n'en est rien. De manière surprenante, le petit phénomène que nous venons de décrire est la seule difficulté à surmonter.

En d'autres termes, nous pouvons officiellement poser :

Définition 2.2 – $\mathbb{R}$ est l'ensemble des écritures décimales avec une infinité de chiffres décimaux après la virgule *qui ne se terminent pas par une infinité de* 9.

On décide donc de bannir une fois pour toutes de l'ensemble $\mathbb{R}$ les écritures pathologiques se terminant par des 9. Si d'aventure quelqu'un écrit un nombre comme $0,99999\ldots$ nous prendrons ça comme une notation pour le nombre 1 (en voyant $0,99999\ldots$ nous devons réagir comme en voyant $12-11$ ou $\frac{5}{5}$).

Avec cette nouvelle définition (correcte), il est possible de montrer le théorème que voici.

Théorème 2.3 – *L'ensemble $\mathbb{R}$ a les propriétés suivantes.*

1. Propriétés arithmétiques. $\mathbb{R}$ *possède une addition et une multiplication telles que :*
   (a) $x + y = y + x$,
   (b) $0 + x = x$,
   (c) $(x+y)+z = x+(y+z)$,
   (d) *pour chaque x il existe un nombre noté $(-x)$ tel que* $x + (-x) = 0$,
   (e) $xy = yx$,
   (f) $1x = x$,
   (g) $(xy)z = x(yz)$,
   (h) $x(y+z) = xy + xz$.
   (i) *pour tout $y \neq 0$ il existe un nombre noté y^{-1} ou $\frac{1}{y}$ tel que* $yy^{-1} = 1$.
2. Propriétés d'ordre. *Les éléments de $\mathbb{R}$ peuvent être comparés. Plus précisément, il y a une relation notée $\leq$ telle que :*
   (a) *étant donnés x et y dans $\mathbb{R}$, on a soit $x \leq y$, soit $y \leq x$,*
   (b) *pour tout x on a $x \leq x$,*
   (c) *si $x \leq y$ et si $y \leq z$, alors $x \leq z$,*
   (d) *si $x \leq y$ et $y \leq x$ alors $x = y$,*
   (e) *si $x \leq y$ alors $x + z \leq y + z$,*
   (f) *si $x \leq y$ et si $0 \leq z$, alors $xz \leq yz$.*
3. Relation avec $\mathbb{Q}$. *On a $\mathbb{Q} \subset \mathbb{R}$, et les opérations usuelles d'addition, de multiplication et d'ordre dans $\mathbb{Q}$ coïncident avec celles calculées dans $\mathbb{R}$. De plus, le nombre décimal*
   $$n, a_1 a_2 \ldots a_k \in \mathbb{Q}$$
   peut être identifié avec le réel
   $$n, a_1 a_2 \ldots a_k 0000\ldots \in \mathbb{R}.$$
   Enfin, si $p, q \in \mathbb{Z}$ avec $q \neq 0$, l'écriture décimale de $\frac{p}{q}$ peut être obtenue par l'algorithme classique, décrit au cours élémentaire.
4. Densité. *Soit $r \geq 0$ un élément de $\mathbb{R}$, et soit $r_n \in \mathbb{Q}$ le décimal obtenu en ne gardant que les n premières décimales de r. Alors*
   $$r_n \leq r < r_n + 10^{-n}.$$
   On en déduit que pour tous a, b dans $\mathbb{R}$ tels que $a < b$, il existe $x \in \mathbb{Q}$ tel que $a < x < b$.

Nous allons commenter ce théorème point par point. Mais d'abord, précisons que nous n'allons pas donner de démonstration – c'est le seul théorème de ce livre que nous ne démontrerons pas, même si on peut indiquer les grandes lignes. L'idée principale est que l'on sait additionner et multiplier les nombres décimaux, et il s'agit d'adapter cette procédure dans le cas d'un nombre infini de décimales. Les difficultés que l'on rencontre sont :

- ◇ Depuis l'école primaire, vous savez additionner et multiplier les nombres décimaux par un algorithme. Remarquez que, dans ces calculs, on commence tout le temps par le chiffre de droite… Que faire quand il y a une infinité de chiffres ? En gros, il faut montrer que l'on peut tronquer les nombres en ne gardant qu'un nombre fini de décimales, faire son opération, et alors les premières décimales que l'on obtient sont correctes.
- ◇ Le problème des « infinités de 9 » va clairement compliquer les choses.
- ◇ Notre discussion des écritures décimales a été bien trop superficielle : il faut commencer par démontrer une quantité de résultats basiques, comme l'unicité par exemple (c'est-à-dire le fait que si $n, a_1 a_2 \ldots a_k = m, b_1 b_2 \ldots b_\ell$ alors $n = m$, $k = \ell$ et $a_i = b_i$ pour $1 \leq i \leq k$).
- ◇ Après pas mal de travail pour montrer que l'addition et la multiplication sont bien définies, on n'a pas encore commencé à montrer les propriétés annoncées dans le théorème …

Tout ceci serait très long, mais finalement sans grande surprise.

Ce système de nombres appelé $\mathbb{R}$ était inconnu des Grecs, même s'ils avaient conscience de l'imperfection de $\mathbb{Q}$. Nous allons voir tout au long de ce livre que les nombres « réels » (les éléments de $\mathbb{R}$) sont parfaitement adaptés à la description du monde réel : ils peuvent mesurer les longueurs et les aires sans mener à des contradictions, par exemple. Nous allons commencer par montrer que 2 possède une racine carrée dans $\mathbb{R}$, bien sûr.

Avant ça, voici quelques remarques supplémentaires :

1. Les propriétés arithmétiques vous sont familières. Il n'y a rien à retenir vraiment, puisque vous les appliqueriez sans réfléchir. Noter que l'on a écrit 0 pour $0,00000\ldots$ et 1 pour $1,00000\ldots$. Nous ferons pareil avec tous les nombres entiers.
2. Même remarque avec les propriétés d'ordre.
3. Dans la troisième série de propriétés, on indique bien que $\mathbb{R}$ contient $\mathbb{Q}$, et pas seulement les nombres décimaux. Par exemple $\frac{1}{3}$ n'est pas un nombre décimal, mais c'est bien un élément de $\mathbb{R}$, et en fait c'est $0,33333\ldots$
4. On dit parfois que « $\mathbb{Q}$ est dense dans $\mathbb{R}$ » pour exprimer le fait qu'entre deux réels a et b, aussi proches que l'on veut, il y aura toujours un rationnel x.

   On peut se demander s'il y a une sorte de « réciproque » au premier énoncé de ce point 4, qui dirait : si x et y sont tels que $x \leq y < x + 10^{-n}$, alors x et y ont « un certain nombre » (soyons prudents) de décimales en commun. C'est tout-à-fait faux : pour $y = 0,999\ldots 9000\ldots$ (prenez un nombre de 9 plus grand que n) et $x = 1$, on a bien les inégalités mais les décimales de x et y sont bien différentes ! Là encore, le phénomène « des 9 » est le seul qui peut vraiment se produire, mais il faut tout de même faire attention à ce que l'on dit.

Faire des calculs concrets avec $\mathbb{R}$ ne nécessite pas d'entraînement spécial, pour la bonne raison que les calculs concrets se font toujours dans $\mathbb{Q}$: ça dépend un peu de ce qu'on appelle « concret » évidemment, mais dans l'ensemble dès que l'on cherche des valeurs numériques, des décimales, on ne va en calculer qu'un nombre fini et donc rester dans les nombres décimaux. Par contre il va être très utile concrètement de savoir que $\sqrt{2}$ existe, et nous avons besoin de $\mathbb{R}$ pour ça.

Terminons avec les propriétés basiques des réels en montrant la très utile « inégalité triangulaire ». On définit, pour $x \in \mathbb{R}$, sa *valeur absolue* $|x|$ par $|x| = x$ si $x \geq 0$ et $|x| = -x$ sinon. Autrement dit, $|x|$ est le plus grand des deux nombres x et $-x$. On en déduit les inégalités suivantes, extrêmement pratiques.

Lemme 2.4 (Inégalité triangulaire) – *Si a et b sont des réels, on a l'inégalité*

$$|a+b| \leq |a| + |b|.$$

Démonstration. Comme $a \leq |a|$ et pareil pour b, en additionnant on trouve $a + b \leq |a| + |b|$. De même $-a \leq |a|$ et pareil pour b, d'où $-(a+b) \leq |a| + |b|$. Comme $|a+b|$ est soit $a+b$ soit $-(a+b)$, l'inegalité est assurée dans tous les cas. □

Corollaire 2.5 (Deuxième inégalité triangulaire) – *Si a et b sont des réels, on a*

$$\big||a| - |b|\big| \leq |a-b|.$$

Démonstration. En appliquant la première inégalité triangulaire à a et $b-a$, on obtient

$$|b| = |a + (b-a)| \leq |a| + |b-a|,$$

d'où $|b| - |a| \leq |b-a|$. En inversant les rôles de a et b, on obtient $|a| - |b| \leq |a-b|$. Ceci donne bien le résultat puisque $|b-a| = |a-b|$, et $\big||a| - |b|\big| = \pm(|a| - |b|)$. □

Si vous avez essayé de réfléchir à la démonstration du théorème 2.3, vous avez du constater que les arguments concernant les décimales des nombres sont délicats, et même assez pénibles (à établir comme à suivre). Si notre but est de montrer qu'il existe un nombre dans $\mathbb{R}$ dont le carré vaut 2, a-t-on envie de se lancer dans une discussion des décimales possibles du nombre ? Même si cela fonctionnait, devra-t-on recommencer de zéro pour $\sqrt{3}$ et $\sqrt{5}$ et tous les autres ? Il paraît douteux qu'on arrive à dégager une méthode générale par une approche calculatoire basée sur les décimales.

Nous allons procéder différemment, et c'est un bon exemple de la démarche générale du mathématicien. Nous allons isoler une propriété de $\mathbb{R}$, appelée « propriété de la borne supérieure », qui va être finalement la seule chose à savoir sur $\mathbb{R}$: toutes les fois où un résultat est faux avec $\mathbb{Q}$, mais devient vrai avec $\mathbb{R}$, vous pouvez être sûr que la « borne supérieure » joue un rôle clef. On peut remercier ici, comme souvent, les générations de mathématiciens qui ont dégagé cette notion pour nous. Sans cette observation, on en serait encore à rédiger des arguments fastidieux avec les décimales.

Dans les paragraphes qui suivent, nous donnons la définition d'une « borne supérieure » et explorons le concept dans $\mathbb{Q}$, pour commencer. Sans surprise, nous verrons que tout ne fonctionne pas comme on le souhaiterait. Plus loin nous verrons que tout marche avec $\mathbb{R}$.

Définition 2.6 – Soit $A \subset \mathbb{Q}$.

- Soit $M \in \mathbb{Q}$. On dit que M est un *majorant* de A si $\forall a \in A$, $a \leq M$.
- Soit $M \in \mathbb{Q}$. On dit que M est *le plus grand élément de* A si c'est un majorant de A et si $M \in A$.
- En remplaçant $\leq$ par $\geq$, on obtient les notions de *minorant* et de *plus petit élément*.
- Soit
$$B = \{M \in \mathbb{Q} \mid M \text{ est un majorant de } A\}.$$
Si B possède un *plus petit* élément b, on dit que c'est la *borne supérieure de* A et on note $b = \sup A$.
- De même, si l'ensemble des minorants de A possède un *plus grand* élément, celui-ci est appelé la *borne inférieure* de A, notée $\inf A$.

On retient que « le sup est le plus petit des majorants », de même que « l'inf est le plus grand des minorants ». Nous allons voir que le sup et l'inf n'existent pas toujours, et c'est bien là le problème. Voyons quelques exemples.

Exemple 2.7 – Soit
$$A = \{x \in \mathbb{Q} \mid 0 \leq x < 1\}.$$
Les minorants de A, pour commencer, sont tous les nombres m tels que $m \leq 0$, c'est-à-dire qu'ils forment l'ensemble
$$C = \{m \in \mathbb{Q} \mid m \leq 0\}.$$
Cet ensemble possède un plus grand élement, à savoir 0. C'est donc le plus grand minorant de A, et par définition on peut écrire $\inf A = 0$. Ce nombre est également le plus petit élément de A.

Nous affirmons que l'ensemble des majorants de A est
$$B = \{M \in \mathbb{Q} \mid M \geq 1\}.$$
Montrons-le. Il est clair que les éléments de B sont des majorants de A, et il faut montrer qu'il n'y en a pas d'autres. Soit donc M un majorant quelconque, et supposons par l'absurde que $M < 1$. On a $M \geq 0$ puisque $0 \in A$, donc $0 \leq M < 1$. Considérons alors $a = \frac{1}{2}(M+1)$. On a $M < a < 1$, donc ce nombre s'est glissé entre M et 1, ce qui est absurde : on a $a \in A$ donc on devrait avoir $a \leq M$. Ainsi $M \geq 1$ comme on souhaitait le montrer.

L'ensemble B possède un plus petit élément, à savoir 1. C'est le plus petit majorant de A, de sorte que $\sup A = 1$. Par contre A n'a pas de plus grand élément.

Les bornes inférieure et supérieure de A sont donc 0 et 1 respectivement, et nous voyons sur cet exemple qu'il s'agit bien des « bornes » naturelles de A au sens intuitif. La différence $\sup A - \inf A = 1 - 0 = 1$ donne une mesure de la taille de A.

Exemple 2.8 – Soit maintenant
$$A = \{x \in \mathbb{Q} \mid x^2 \leq 2\}.$$
Intéressons-nous aux majorants de A, et notons comme d'habitude B l'ensemble qu'ils forment. Cet ensemble est non-vide : on a par exemple $10 \in B$ puisque tous les éléments de A sont ≤ 10. En effet, un nombre $x > 10$ satisfait $x^2 > 10^2 = 100 > 2$ et ne peut pas être dans A.

Pour les mêmes raisons, on a $3 \in B$ puisque $3^2 = 9 > 2$. Approchons-nous encore : on voit que $\frac{3}{2} \in B$ puisque $(\frac{3}{2})^2 = \frac{9}{4} > 2$.

Bien. Supposons que B possède un plus petit élément ℓ ; en d'autres termes, supposons que A possède une borne supérieure. Que peut-on dire de ℓ^2 ? En particulier, ce nombre est-il plus grand ou plus petit que 2 ?

Examinons l'éventualité $\ell^2 > 2$. Notons $\varepsilon = \ell^2 - 2 > 0$, et prenons $\delta = \frac{\varepsilon}{2\ell}$. Si on calcule
$$(\ell - \delta)^2 = \ell^2 + \delta^2 - 2\delta\ell,$$
on s'aperçoit de la chose suivante : l'inégalité
$$2\delta\ell - \delta^2 < 2\delta\ell = \varepsilon = \ell^2 - 2$$
entraîne $(\ell - \delta)^2 > \ell^2 + (2 - \ell^2) = 2$. Le nombre $M = \ell - \delta$ est donc un majorant de A puisque son carré est > 2, par le même raisonnement qui nous a servi à montrer que 10, 3 et $\frac{3}{2}$ sont des majorants.

Mais c'est absurde puisque $M < \ell$ et que ℓ est censé être le plus petit majorant ! Cette contradiction réfute l'hypothèse selon laquelle $\ell^2 > 2$, et on en tire $\ell^2 \leq 2$.

On peut maintenant se demander si $\ell^2 < 2$. Dans cette hypothèse, notons $\varepsilon = 2 - \ell^2 > 0$. Choisissons n'importe quel nombre $\delta > 0$ tel que l'on ait à la fois $\delta < 2\ell$ et $\delta < \frac{\varepsilon}{4\ell}$. On note alors que $\delta^2 < 2\delta\ell$ et donc que
$$\delta^2 + 2\delta\ell < 4\delta\ell < \varepsilon.$$
Par suite, $(\ell + \delta)^2 < \ell^2 + \varepsilon = 2$. Donc $a = \ell + \delta$ est un élément de A. C'est de nouveau absurde puisque $a > \ell$ alors que ℓ est un majorant.

Il ne nous reste pas d'autre choix que d'envisager que $\ell^2 = 2$. Mais c'est également impossible en vertu de la proposition 2.1 !

Finalement, cette borne supérieure ℓ ne pourrait satisfaire ni $\ell^2 < 2$, ni $\ell^2 > 2$, ni $\ell^2 = 2$. On en arrive à la conclusion que *l'ensemble* A *ne possède pas de borne supérieure*.

En remplaçant $\mathbb{Q}$ par $\mathbb{R}$ dans la définition 2.6, on obtient les notions de bornes supérieures et inférieures dans $\mathbb{R}$. Or, on a maintenant le théorème essentiel suivant.

THÉORÈME 2.9 – *Soit* $A \subset \mathbb{R}$ *une partie non-vide et majorée. Alors* A *possède une borne supérieure dans* $\mathbb{R}$.

Une partie $A \subset \mathbb{R}$ *non-vide et minorée, de même, possède une borne inférieure dans* $\mathbb{R}$.

La démonstration est un peu compliquée, puisqu'on raisonne sur les décimales (que faire d'autre à ce stade ?), mais ça sera la dernière fois que nous aurons à le faire !

Démonstration. On va montrer l'existence de la borne supérieure, l'autre cas étant similaire. Pour faciliter les notations, on va supposer que A possède au moins un élement ≥ 0, et donc que tous les majorants de A sont ≥ 0 (exercice : comment faire pour se ramener toujours à ce cas ?). La borne supérieure va s'écrire $x = n, d_1 d_2 \ldots$ et nous allons trouver la partie entière $n \in \mathbb{N}$ et les décimales d_i petit à petit.

Soit $B \subset \mathbb{N}$ l'ensemble de nombres entiers N qui sont des majorants de A. L'ensemble B est non-vide, il possède donc un plus petit élément N_0. Nous allons suivre l'argument sur un exemple (en parallèle du cas général), donc disons que $N_0 = 5$. Si $N_0 \in A$, alors c'est le plus grand élement de A (puisque c'est un majorant), c'est donc le sup et on a fini. Supposons donc que $N_0 \notin A$. Le nombre $n = N_0 - 1$, donc $n = 4$ dans l'exemple, est la partie entière que nous cherchons (on va le montrer).

Par définition n n'est pas un majorant de A, mais $n+1$ si, donc il existe des éléments de A qui s'écrivent $n, a_1 a_2 \ldots$. Soit c le plus grand chiffre en première position après la virgule parmi les éléments de A plus grands que n ; si $c = 7$ par exemple, cela signifie que A contient un ou plusieurs éléments de la forme $4,7\ldots$ mais aucun de la forme $4,8\ldots$ Si $c < 9$ on pose $c' = c + 1$ et $x_1 = n, c'$ (dans l'exemple $x_1 = 4,8$) ; alors x_1 est un majorant de A, et si $x_1 \in A$ alors $x_1 = \sup A$ et on a fini (donc $d_1 = c'$ et $d_i = 0$ pour $i \geq 2$). Sinon, si $x_1 \notin A$, on pose $d_1 = c$ et on continue. Dans le cas où $c = 9$ on pose $x_1 = n$ et $d_1 = 9$ et on poursuit également.

Maintenant on redéfinit c comme étant le plus grand chiffre en deuxième position après la virgule parmi les éléments de A plus grands que n, d_1 ; si par exemple $c = 5$, alors A contient des éléments qui s'écrivent $4,75\ldots$ mais aucun qui s'écrit $4,76\ldots$ Si le majorant $x_2 = 4,76$ appartient à A, c'est le sup et on a fini ; sinon on continue, en posant $d_2 = 5$.

Un dernier tour explicite : on définit c comme étant le plus grand chiffre en troisième position après la virgule parmi les éléments de A plus grands que $4,75$; admettons que dans l'exemple on ait cette fois-ci $c = 9$, donc A possède des élements de la forme $4,759\ldots$ (et dans le cas d'un 9, on sait déjà depuis l'étape précédente que A ne possède pas d'élément qui commence par $4,76\ldots$). On pose $d_3 = 9$ et $x_3 = x_2 = 4,76$.

On va continuer ce procédé autant que possible. Si à un moment donné l'un des $x_i \in A$, alors comme on l'a dit c'est un sup et la démarche s'arrête. Supposons donc que l'on soit forcé de continuer sans cesse, appliquant l'algorithme ci-dessus qui produit des décimales $d_1, d_2, d_3, \ldots$ et des majorants $x_1, x_2, x_3 \ldots$ de A. On note que x_k est un nombre décimal, qui commence par $n, d_1 d_2 \ldots d_{k-1}$; sa k-ième décimale est soit $d_k + 1$ si $d_k < 9$, soit 0 sinon, alors que toutes les décimales suivantes sont nulles. Par ailleurs si x'_k désigne le nombre obtenu en diminuant de 1 la dernière décimale non-nulle de x_k, alors x'_k n'est pas un majorant de A.

Posons alors $x = n, d_1 d_2 d_3 \ldots$, de sorte que $x \leq x_k$ pour tout k. Si on peut montrer que x est un majorant de A, alors ce sera visiblement le sup : en effet tout nombre $< x$ doit être $\leq x'_k$ pour un certain k, et donc n'est pas un majorant.

Supposons donc par l'absurde qu'il existe $a \in A$ tel que $a > x$. D'après le (4) du théorème 2.3 on a $r_k \leq x < r_k + 10^{-k}$, où r_k est le nombre obtenu à partir de x en tronquant à la k-ième décimale ; de même, puisque x et x_{k+1} ont k décimales en commun, on a $r_k \leq x_{k+1} < r_k + 10^{-k}$. On en déduit en soustrayant que $-10^{-k} \leq x_{k+1} - x \leq 10^{-k}$, et donc $x_{k+1} \leq x + 10^{-k}$. En prenant k suffisamment grand, on aura $x + 10^{-k} < a$, d'où l'absurde conclusion que $x_{k+1} < a$ alors que x_{k+1} est un majorant. □

Ce théorème est la seule chose à savoir sur $\mathbb{R}$, vraiment. Nous le disions ci-dessus, nous avons laissé pas mal de détails de côté en donnant la définition de $\mathbb{R}$, mais ça n'a aucune importance pour la suite, si vous savez que les sup et les inf existent dans $\mathbb{R}$. Par ailleurs, si vous comparez avec d'autres manuels (ou avec la page Wikipedia), vous verrez que l'on définit parfois l'ensemble des réels bien différemment (certains utilisent les « suites de Cauchy », d'autres les « coupures de Dedekind »...). Le fait même que plusieurs constructions existent montre bien que les particularités de l'une d'entre elles ne sont pas cruciales. Tant que l'on se met d'accord sur l'énoncé du théorème 2.3 et celui du théorème 2.9, on parle nécessairement de la même chose (voir par ailleurs l'encadré « Unicité de $\mathbb{R}$ »).

Unicité de $\mathbb{R}$

Soit $\mathbb{R}'$ un ensemble pour lequel les énoncés des théorèmes 2.3 et 2.9 sont valides. Alors on peut montrer que $\mathbb{R}'$ peut être identifié avec $\mathbb{R}$, et voici une toute petite esquisse de démonstration.

Soit $x \in \mathbb{R}$, et soit $A_x = \{q \in \mathbb{Q} \mid q \leq x\}$. On peut montrer facilement que $\sup A_x = x$, en prenant le sup dans $\mathbb{R}$ (faites-le). Mais $\mathbb{Q} \subset \mathbb{R}'$ donc $A_x \subset \mathbb{R}'$, et on peut prendre aussi le sup dans $\mathbb{R}'$ puisque $\mathbb{R}'$ vérifie le théorème 2.9 par hypothèse. Soit donc $x' = \sup A_x$, le sup étant pris dans $\mathbb{R}'$.

La règle $x \mapsto x'$ que nous venons de définir donne une fonction $\mathbb{R} \to \mathbb{R}'$, et on peut montrer que c'est une bijection. De plus cette bijection est compatible avec l'addition, la multiplication, la relation d'ordre $\leq$, et $x' = x$ lorsque $x \in \mathbb{Q}$. On peut donc sans risque identifier $\mathbb{R}'$ et $\mathbb{R}$.

voir les exercices 477, 470, 472

En guise de première application, voici enfin le résultat qui indique que les racines carrées existent dans $\mathbb{R}$.

PROPOSITION 2.10 – *Soit* $a \in \mathbb{R}$ *un nombre positif, c'est-à-dire* $a \geq 0$. *Alors il existe un nombre* $x \in \mathbb{R}$, *et un seul, tel que* $x \geq 0$ *et* $x^2 = a$.

On note ce nombre $\sqrt{a}$ *et on l'appelle la racine carrée de* a.

Démonstration. Cette démonstration va nous donner l'occasion unique d'indiquer à plusieurs reprises quelles propriétés du théorème 2.3 nous sont utiles, même les plus évidentes, pour que vous voyiez bien que les raisonnements habituels reposent toujours sur ces quelques règles (nous n'avons pas encore mis l'accent là-dessus, et nous ne le ferons plus par la suite). On ne va pas indiquer *toutes* les propriétés que l'on utilise – ça serait long et pénible à suivre – mais seulement quelques unes.

Voyons d'abord l'unicité. Soit donc $x_0 \geq 0$ tel que $x_0^2 = a$, et cherchons tous les $x \geq 0$ tels que $x^2 = a$. On a

$$\begin{aligned} x^2 = a &\iff x^2 - x_0^2 = 0, \\ &\iff (x - x_0)(x + x_0) = 0. \end{aligned}$$

(Pourquoi cette factorisation est-elle valide ?) Lorsqu'un produit ab de deux nombres réels a et b vaut 0, l'un de ces nombres doit être nul : en effet si $a \neq 0$, alors $\frac{1}{a}$ existe d'après le (1)(i) du théorème 2.3, et en multipliant par $\frac{1}{a}$ on obtient $\frac{1}{a}(ab) = (\frac{1}{a}a)b = 1b = b = \frac{1}{a}0 = 0$ (nous avons utilisé les propriétés (1)(g), puis (1)(i), puis (1)(f) ; le fait que $x0 = 0$ pour tout x se montre à partir des propriétés : faites-le !). Donc $b = 0$.

Ici (avec $a = x - x_0$ et $b = x + x_0$) on voit que $x = x_0$ ou $x = -x_0$. Comme $0 \leq x_0$, on observe en ajoutant $-x_0$ de chaque côté que $-x_0 \leq 0$ (propriété (2)(e) du théorème). Ainsi, dans le cas où $x = -x_0$, on constate que x est à la fois ≤ 0 et ≥ 0 ; c'est donc que $x = 0$ par la propriété (2)(d), d'où $x = x_0 = 0$. Finalement $x = x_0$ quoi qu'il arrive, et l'unicité est démontrée.

Passons à l'existence. Sans surprise, on pose

$$A = \{x \in \mathbb{R} \mid x^2 \leq a\}.$$

Si l'on trouve un nombre M tel que $M^2 > a$, alors ce sera un majorant de A. (Pour vérifier ceci, démontrez que si $x \geq M$, alors $x^2 \geq M^2$ lorsque $M \geq 0$.) Or on peut tout simplement prendre $M = a$ si $a > 1$, et $M = 1$ si $a < 1$ (pour le cas $a = 1$, la proposition est évidemment vraie).

Puisque l'on est dans $\mathbb{R}$, on sait que A possède une borne supérieure, en tant qu'ensemble non-vide (il contient 0) et majoré ; posons donc $x = \sup A$. Il faut montrer que $x^2 = a$. Nous avons déjà fait le raisonnement dans l'exemple 2.8 dans le cas où $a = 2$; en procédant exactement de la même manière, on trouve que $x^2 < a$ et $x^2 > a$ mènent à des contradictions. Donc $x^2 = a$. □

Par la suite, nous montrerons même que tout nombre positif a possède une unique « racine n-ième » notée $\sqrt[n]{a}$ ou $a^{\frac{1}{n}}$, c'est-à-dire qu'il existe un unique nombre positif x tel que $x^n = a$. Vous pouvez essayer de démontrer ce résultat maintenant sur le même modèle : c'est un peu fastidieux, mais on y arrive. Nous allons préférer déduire le résultat du très utile « théorème des valeurs intermédiaires » (4.8), qui montrera tout son intérêt.

voir les exercices 456, 461

Nous n'avons pas encore toutes les racines carrées que l'on pourrait souhaiter : il manque encore les racines des nombres négatifs. En effet si $x \in \mathbb{R}$, alors $x^2 \geq 0$, et donc par exemple il n'y a pas de nombre réel dont le carré serait -1. Il nous faut donc un système de nombres encore plus étendu.

Cette fois-ci, les choses sont beaucoup plus simples. Nous allons voir qu'il suffit de « rajouter » un nombre i tel que $i^2 = -1$, et toutes les racines carrées imaginables sont obtenues – et même bien plus.

Comment donc « rajouter » ce i ? Si notre nouveau système de nombres contient $\mathbb{R}$ et un tel nombre i, alors il doit contenir des nombres de la forme $x + iy$ avec $x, y \in \mathbb{R}$. De plus si les règles usuelles de calcul s'appliquent (ce que l'on souhaite), on doit avoir

$$(x+iy)+(x'+iy') = (x+x')+i(y+y'),$$

ainsi que

$$(x+iy)(x'+iy') = (xx'-yy')+i(xy'+x'y).$$

Ceci motive la définition suivante.

Définition 2.11 – Sur le produit cartésien $\mathbb{R}\times\mathbb{R}$, on définit une addition par

$$(x,y)+(x',y') = (x+x', y+y'),$$

et une multiplication par

$$(x,y)(x',y') = (xx'-yy', xy'+x'y).$$

Muni de ces deux opérations, l'ensemble $\mathbb{R}\times\mathbb{R}$ est noté $\mathbb{C}$, et appelé ensemble des nombres complexes.

Voyons pourquoi cette définition est la bonne. Si x, x' sont réels, on a

$$(x,0)+(x',0) = (x+x',0) \quad \text{et} \quad (x,0)(x',0) = (xx',0).$$

Donc l'ensemble des nombres complexes de la forme $(x,0)$ se comporte exactement comme l'ensemble $\mathbb{R}$. On peut identifier ces deux ensembles sans risque de confusion, et lorsque $x \in \mathbb{R}$ on écrira également x pour le nombre complexe $(x,0)$. (Le lecteur qui a pris connaissance de la définition 1.13 parlera plutôt d'une bijection $x \mapsto (x,0)$, qui se trouve être compatible avec les opérations arithmétiques.)

Ensuite, posons $i = (0,1)$. On a bien

$$\begin{aligned} i^2 &= (0,1)(0,1) \\ &= (-1,0) \\ &= -1. \end{aligned}$$

Enfin, pour tout réel y, on a $iy = (0,1)(y,0) = (0,y)$. Finalement tout nombre complexe (x,y) peut s'écrire $(x,y) = (x,0)+(0,y) = x+iy$.

On vient de montrer que $\mathbb{R} \subset \mathbb{C}$, que $\mathbb{C}$ contient une racine de -1, et visiblement $\mathbb{C}$ ne pouvait pas être plus petit. Tout se passe décidément bien, puisqu'on a le résultat suivant :

Proposition 2.12 – *L'ensemble $\mathbb{C}$ satisfait les neuf propriétés (1) (a-b-c-d-e-f-g-h-i) du théorème 2.3. En d'autres termes, les règles de calcul usuelles s'appliquent.*

Démonstration. Ce sont des vérifications évidentes, sauf pour la (1)(i). Étant donné $z = x+iy$ un nombre complexe non-nul, il faut trouver un nombre complexe w tel que $zw = 1$. Un tel nombre, s'il existe, serait évidemment unique, et ce serait z^{-1}.

On appelle *conjugué* de z le nombre $\overline{z} = x - iy$. On a $z\overline{z} = x^2+y^2$; ce dernier nombre est un réel positif, on peut noter $|z| = \sqrt{x^2+y^2}$, que l'on appelle le *module* de z. Notons que lorsque $z \neq 0$, on a $|z| > 0$; en particulier $\frac{1}{|z|}$ existe.

Soit alors

$$w = \frac{\overline{z}}{|z|^2} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}.$$

C'est bien un nombre complexe, et on a $zw = \frac{z\overline{z}}{z\overline{z}} = 1$. Donc $w = z^{-1}$. □

Remarque 2.13. L'opération de conjugaison que nous venons d'utiliser possède de bonnes propriétés : en effet $\overline{z+w} = \overline{z}+\overline{w}$ et $\overline{zw} = \overline{z}\,\overline{w}$ pour tout $z, w \in \mathbb{C}$, comme on le vérifie facilement. Par suite on a

$$|zw|^2 = zw\overline{zw} = z\overline{z}w\overline{w} = |z|^2|w|^2.$$

En prenant les racines carrées de ces nombres réels, on a également $|zw| = |z||w|$. *voir les exercices 1, 5127*

Nul besoin de chercher de système de nombres plus grand pour obtenir des racines carrées. En effet :

Proposition 2.14 – *Tout nombre complexe $w \in \mathbb{C}$ non-nul possède exactement deux racines carrées, qui sont opposées.*

« Opposées » signifie que l'on a deux racines z et $-z$, et aucune autre.

Démonstration. Il est très facile de voir que si w possède une racine z_0, alors il en possède exactement deux : en effet

$$\begin{aligned} z^2 = w &\iff z^2 - z_0^2 = 0 \\ &\iff (z-z_0)(z+z_0) = 0 \\ &\iff z = z_0 \text{ ou } z = -z_0. \end{aligned}$$

Pour l'existence, écrivons $w = a+ib$, et cherchons un nombre $z = x+iy$ tel que $z^2 = w$. Ceci revient à résoudre

$$\begin{cases} x^2 - y^2 = a & (1) \\ 2xy = b & (2). \end{cases}$$

Il est astucieux ici de regarder les modules : on doit avoir $|z|^2 = |z^2| = |w|$, et donc

$$x^2 + y^2 = \sqrt{a^2+b^2} \qquad (3).$$

En faisant $(1)+(3)$ on tire

$$x^2 = \frac{a+\sqrt{a^2+b^2}}{2}.$$

Le membre de droite est un réel ≥ 0, donc cette dernière équation a bien des solutions, ce qui donne deux choix opposés pour x. De même en faisant $(3)-(1)$ on obtient

$$y^2 = \frac{-a+\sqrt{a^2+b^2}}{2}.$$

Là encore on a deux possibilités pour y.

Quels que soient les choix, l'équation (1) est satisfaite ; quant à l'équation (2), on est seulement assuré d'avoir $4x^2y^2 = b^2$ et donc $2xy = \pm b$. Il suffit alors d'ajuster le signe de x ou de y pour satisfaire (2). □

Exemple 2.15 – Soit $w = 1-2i$, et cherchons $z = x+iy$ tel que $z^2 = w$. Comme dans la démonstration, on constate que l'on doit avoir

$$\begin{cases} x^2 - y^2 = 1 \\ 2xy = -2 \\ x^2+y^2 = \sqrt{5}. \end{cases}$$

Toujours en suivant le modèle de la démonstration, on en déduit

$$x^2 = \frac{1+\sqrt{5}}{2} \quad \text{et} \quad y^2 = \frac{-1+\sqrt{5}}{2}.$$

L'équation $2xy = -2$ nous dit que x et y doivent être de signes opposés. On peut donc prendre

$$x = \sqrt{\frac{1+\sqrt{5}}{2}} \quad \text{et} \quad y = -\sqrt{\frac{-1+\sqrt{5}}{2}}.$$

Les deux solutions sont alors $x+iy$ et $-x-iy$.

On sait même résoudre dans $\mathbb{C}$ des équations un peu plus compliquées :

Proposition 2.16 – *Soient a, b et c trois nombres complexes, soit $\Delta = b^2 - 4ac$, et enfin soit $\delta \in \mathbb{C}$ tel que $\delta^2 = \Delta$. On suppose que $a \neq 0$.*

Alors l'équation

$$az^2 + bz + c = 0$$

possède exactement deux solutions lorsque $\Delta \neq 0$, données par :

$$z = \frac{-b \pm \delta}{2a}.$$

Dans le cas où $\Delta = 0$, ces deux solutions se confondent et il n'y en a pas d'autres.

Démonstration. On écrit simplement

$$\begin{aligned} az^2+bz+c = 0 &\iff 4a^2z^2 + 4abz + 4ac = 0 \\ &\iff (2az+b)^2 + 4ac - b^2 = 0 \\ &\iff (2az+b)^2 = \Delta, \end{aligned}$$

et donc $2az+b$ doit être $\pm\delta$. □ *voir les exercices 27, 31, 2945*

À ce stade, vos souvenirs de Terminale vous poussent sans doute à attendre une description de la « forme polaire », la fameuse écriture $z = \rho e^{i\theta}$. En réalité, pour expliquer rigoureusement ce qu'il se passe, il va falloir patienter : il nous faut d'abord voir une quantité d'autres résultats. En contrepartie, quand nous arriverons enfin à cette écriture, nous aurons des vraies *définitions* du cosinus et du sinus, entre autres choses.

Vous pouvez toutefois aller voir tout de suite le passage consacré aux nombres complexes dans l'appendice A. On y donne (sans justification !) les mêmes explications qu'au lycée.

– Fin de la première lecture –

Les systèmes de nombres $\mathbb{R}$ et $\mathbb{C}$ semblent répondre à tous nos besoins en théorie. En pratique par contre, les choses ne sont pas aussi simples. Dès que l'on commence à faire des calculs un peu longs, le besoin de confier la tâche à un ordinateur se fait sentir. Or, les nombres réels sont très abstraits, nous l'avons dit ; tout ce qu'une machine va savoir faire, c'est utiliser des approximations, comme par exemple

$$x = 1.414213 = \frac{1414213}{1000000}$$

pour approcher $\sqrt{2}$. En fait les machines ne connaissent que $\mathbb{Q}$ (et encore, avec des limitations sur la taille des nombres entiers employés en fonction de la mémoire, mais on peut laisser ce problème de côté). Ces approximations sont une source d'erreur importante. Ainsi si l'on calcule

$$x^{32} = 65535,1660562286\cdots$$

on est bien loin de $\sqrt{2}^{32} = 2^{16} = 65536$. Même en sachant que x^{32} est censé approcher un nombre entier, arrondir à l'entier le plus proche ne donne pas la bonne réponse ! Aussi, notons que 7 chiffres de x sont corrects, alors que seulement 4 chiffres de x^{32} sont corrects.

Cependant, admettons que l'on entreprenne une série de calculs, dans lesquels on est certain de n'utiliser rien d'autre que des nombres rationnels et $\sqrt{2}$. On peut tout simplement apprendre à l'ordinateur à manipuler les nombres de la forme $a + b\sqrt{2}$ avec $a, b \in \mathbb{Q}$. En effet il suffit de stipuler les règles suivantes :

$$(a + b\sqrt{2}) + (a' + b'\sqrt{2}) = (a + a') + (b + b')\sqrt{2},$$

et

$$(a + b\sqrt{2})(a' + b'\sqrt{2}) = (aa' + 2bb') + (ab' + a'b)\sqrt{2}.$$

On conçoit bien comment un ordinateur peut considérer ces nombres commes des paires (a, b) de rationnels et opérer les additions et multiplications directement sur ces paires (un peu comme dans notre définition de $\mathbb{C}$ avec des paires de réels). Voilà un nouveau système de nombres qui apparaît naturellement, et sur le même modèle on en entrevoit une infinité. Il est temps de leur donner des noms précis.

« Anneau » est une mauvaise traduction de l'allemand Ringe, *qui signifie « cercle », dans le sens de « communauté ».*

Définition 2.17 – On dit que l'ensemble $\mathbb{K}$ est un *anneau* lorsqu'il est muni d'une addition

$$\begin{array}{ccc} \mathbb{K}\times\mathbb{K} & \longrightarrow & \mathbb{K} \\ (x,y) & \mapsto & x+y \end{array}$$

et d'une multiplication

$$\begin{array}{ccc} \mathbb{K}\times\mathbb{K} & \longrightarrow & \mathbb{K} \\ (x,y) & \mapsto & x\cdot y \end{array}$$

ainsi que de deux éléments notés 0 et 1, tels que les propriétés suivantes sont satisfaites :

(a) $x + y = y + x$,
(b) $0 + x = x$,
(c) $(x + y) + z = x + (y + z)$,
(d) $\forall x\ \exists(-x)$ tel que $x + (-x) = 0$,
(e) $x(y + z) = xy + xz$
et $(x + y)z = xz + yz$,
(f) $1x = x1 = x$,
(g) $(xy)z = x(yz)$.

Lorsque de plus on a

$$\text{(h)} \quad xy = yx,$$

on dit que $\mathbb{K}$ est un *anneau commutatif*.

Finalement, si en plus des propriétés (a-b-c-d-e-f-g-h) on a également

$$\text{(i)} \quad \forall x \neq 0\ \exists x^{-1} \text{ tel que } xx^{-1} = 1,$$

on dit que $\mathbb{K}$ est un *corps*.

En d'autres termes, dans un corps les règles usuelles de calcul doivent s'appliquer.

L'expression « corps » à l'origine était comprise dans le sens d'un corps de métier, ou d'un corps d'armée.

Exemple 2.18 – Les ensembles $\mathbb{Q}$, $\mathbb{R}$ et $\mathbb{C}$, avec les opérations usuelles, sont des corps.

Exemple 2.19 – Soit $\mathbb{K} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ avec les opérations définies ci-dessus. On va montrer que $\mathbb{K}$ est un corps. En fait les opérations sont « héritées » de celles de $\mathbb{R}$ (notez que l'on a $\mathbb{K} \subset \mathbb{R}$), et par conséquent les propriétés a-b-c-d-e-f-g-h sont automatiquement satisfaites. Mais il en manque une !

En effet, il n'est pas évident que si $x = a + b\sqrt{2} \in \mathbb{K}$, alors $x^{-1} \in \mathbb{K}$ lorsque $x \neq 0$ (on est simplement certain que x^{-1} existe dans $\mathbb{R}$). Cependant un petit calcul nous rassure :

$$\begin{aligned} \frac{1}{x} = \frac{1}{a + b\sqrt{2}} &= \frac{a - b\sqrt{2}}{(a + b\sqrt{2})(a - b\sqrt{2})} \\ &= \frac{a}{a^2 - 2b^2} + \frac{-b}{a^2 - 2b^2}\sqrt{2} \in \mathbb{K}. \end{aligned}$$

Un avertissement. Nous avons multiplié numérateur et dénominateur par $a - b\sqrt{2}$, et ceci n'a un sens que si $a - b\sqrt{2} \neq 0$. Or dans le cas contraire, on aurait $\sqrt{2} = \frac{a}{b}$ avec a et b rationnels, ce qui est impossible d'après la proposition 2.1.

Ce corps est noté généralement $\mathbb{Q}[\sqrt{2}]$.

Exemple 2.20 – L'ensemble $\mathbb{Z}$ est un anneau commutatif, mais ce n'est pas un corps. En effet la propriété (i) de l'inverse n'est pas satisfaite, par exemple $\frac{1}{2} \notin \mathbb{Z}$.

Il va falloir attendre le chapitre 12 pour avoir un exemple d'anneau non-commutatif.

Nous allons donner d'autres exemples de corps, qui ne possèdent qu'un nombre fini d'éléments. Ils sont utilisés extrêmement souvent en théorie des nombres, en informatique, en cryptographie, etc.

L'idée de départ est simple. Lorsqu'il est 23h et qu'on attend un événement qui doit se dérouler 4h plus tard, on calcule rapidement qu'il aura lieu à 3h du matin. S'il est 19h et que l'on a 7h à attendre, on sait bien que cela va nous amener à 2h du matin. Le raisonnement que l'on fait sans y penser consiste à additionner les deux nombres (on obtient 27 dans le premier cas, et 26 dans le deuxième), puis à retrancher 24 puisque les journées reprennent à 0 à ce moment-là.

On dit que l'on calcule *modulo* 24. Vous savez aussi spontanément calculer modulo 12 : il suffit de ne pas différencier le matin et l'après-midi, comme lorsqu'on vous demande d'attendre pendant 5h à partir de 11h et que vous savez presque immédiatement que vous en avez jusqu'à 4h (de l'après-midi). Là encore on fait $11+5=16$ puis $16-12=4$ puisque l'on veut un résultat entre 0 et 12.

On va définir maintenant des opérations modulo N, pour tout entier $N \geq 2$, sur le même modèle. Rappelons avant de commencer ce qu'est une *division euclidienne* : étant donnés deux nombres entiers a et $b \neq 0$, vous savez que l'on peut trouver deux nombres entiers q (le quotient) et r (le reste), uniques, tels que

$$a = bq + r\,,$$

et $0 \leq r < b$. Par exemple, en faisant la division de 16 par 12 on écrit $16 = 12\times 1 + 4$ donc $q = 1$ et $r = 4$.

Définition 2.21 – Sur l'ensemble $\{0,1,2,\ldots,N-1\}$ on définit une addition par

$$x \oplus y = \text{ le reste dans la division euclidienne de } x+y \text{ par } N,$$

et une multiplication par

$$x \otimes y = \text{ le reste dans la division euclidienne de } xy \text{ par } N.$$

On écrit $\mathbb{Z}/N\mathbb{Z}$ pour désigner l'ensemble $\{0,1,2,\ldots,N-1\}$ muni de ces opérations.

De plus les éléments de $\mathbb{Z}/N\mathbb{Z}$ vont être notés $\overline{0}, \overline{1}, \ldots, \overline{N-1}$ (pour les distinguer des mêmes éléments vus dans $\mathbb{Z}$).

Exemple 2.22 – Prenons $N = 24$. Si l'on voit $\underline{23}$ et 4 comme des éléments de $\mathbb{Z}/24\mathbb{Z}$, on peut calculer $\overline{23}\oplus\overline{4}$. Comme

$$23+4 = 27 = 24\times 1 + 3\,,$$

on a $\overline{23}\oplus\overline{4} = \overline{3}$. De même :

$$23\times 4 = 92 = 24\times 3 + 20\,,$$

donc $\overline{23}\otimes\overline{4} = \overline{20}$.

Exemple 2.23 – Prenons $N = 2$; on a maintenant $\mathbb{Z}/2\mathbb{Z} = \{\overline{0},\overline{1}\}$. Le seul calcul un peu étonnant est $\overline{1}\oplus\overline{1} = \overline{0}$: en effet

$$1+1 = 2 = 2\times 1 + 0\,.$$

Sinon on a sans surprise $\overline{0}\oplus\overline{0} = \overline{0}$, et $\overline{0}\oplus\overline{1} = \overline{1}\oplus\overline{0} = \overline{1}$. La multiplication ne vous étonnera pas non plus. Écrivons les résultats complets sous forme de tableaux (sans les « barres » pour alléger) :

⊕	0	1
0	0	1
1	1	0

et

⊗	0	1
0	0	0
1	0	1

Peut-on appliquer les règles de calcul usuelles avec $\mathbb{Z}/N\mathbb{Z}$? Pour vérifier ceci, définissons la fonction « reste » :

$$\begin{array}{rcl} R : \mathbb{Z} & \longrightarrow & \mathbb{Z}/N\mathbb{Z} \\ x & \mapsto & R(x) = \text{le reste dans la division de } x \text{ par } N. \end{array}$$

Proposition 2.24 – *On a*

$$R(x+y) = R(x)\oplus R(y) \quad \text{et} \quad R(xy) = R(x)\otimes R(y).$$

Démonstration. Écrivons les divisions euclidiennes $x = Nq_1 + R(x)$ et $y = Nq_2 + R(y)$. En additionnant on trouve

$$x + y = N(q_1+q_2) + (R(x)+R(y)).$$

Il se peut que $R(x)+R(y)$ soit $\geq N$; écrivons donc une nouvelle division

$$R(x) + R(y) = Nq_3 + r.$$

Ici par définition le reste $r = R(x)\oplus R(y)$. En regroupant :

$$x + y = N(q_1+q_2+q_3) + r\,,$$

et $0 \leq r < N$, donc r est bien le reste dans la division euclidienne de $x+y$ par N. C'est-à-dire que $r = R(x+y) = R(x)\oplus R(y)$.

On procède de même pour la multiplication. □

Corollaire 2.25 – *Les propriétés de calcul (a-b-c-d-e-f-g-h) sont valables dans* $\mathbb{Z}/N\mathbb{Z}$*, pour l'addition* $\oplus$ *et la multiplication* $\otimes$*. Les rôles de 0 et 1 sont joués par* $\overline{0}$ *et* $\overline{1}$*.*

(En d'autres termes, $\mathbb{Z}/N\mathbb{Z}$ *est un anneau commutatif.)*

Démonstration. Soient x, y et z des entiers. Puisque $x(yz) = (xy)z$, en appliquant la fonction R on a $R(x(yz)) = R((xy)z)$, ce qui donne en utilisant la proposition

$$R(x)\otimes R(yz) = R(xy)\otimes R(z),$$

puis

$$R(x)\otimes(R(y)\otimes R(z)) = (R(x)\otimes R(y))\otimes R(z).$$

Par suite, la multiplication $\otimes$ est associative (propriété (g)). On fait pareil pour les autres propriétés. □

L'ensemble $\mathbb{Z}/N\mathbb{Z}$ est-il un corps ? En d'autres termes, que peut-on dire de la règle (i) de l'inverse ? Voyons des exemples.

Exemple 2.26 – Pour $N = 2$, le seul élément non-nul de $\mathbb{Z}/2\mathbb{Z}$ est $x = \overline{1}$. On a $x\otimes x = \overline{1}$, donc x^{-1} existe et on a même $x^{-1} = x$. Ainsi $\mathbb{Z}/2\mathbb{Z}$ est un corps.

Exemple 2.27 – Prenons $N = 24$. Comme

$$3\times 8 = 24 = 24\times 1 + 0\,,$$

on a $\overline{3}\otimes\overline{8} = \overline{0}$. On en déduit que l'inverse de $\overline{3}$ n'existe pas : en effet si l'on avait un élément $(\overline{3})^{-1}$ tel que $(\overline{3})^{-1}\otimes\overline{3} = \overline{1}$, on aurait

$$(\overline{3})^{-1}\otimes(\overline{3}\otimes\overline{8}) = ((\overline{3})^{-1}\otimes\overline{3})\otimes\overline{8} = \overline{1}\otimes\overline{8} = \overline{8} = (\overline{3})^{-1}\otimes\overline{0} = \overline{0}.$$

Or $\overline{8}\neq\overline{0}$. Donc $\mathbb{Z}/24\mathbb{Z}$ n'est pas un corps.

Dans le chapitre 11 nous allons déterminer les valeurs de N telles que $\mathbb{Z}/N\mathbb{Z}$ est un corps. Vous pouvez essayer de deviner la réponse.

Nous allons conclure ce chapitre par une simplification des notations. Il est clair qu'écrire $x\oplus y$ et $x\otimes y$ va devenir fatigant très vite, donc on va noter $x+y$ et xy. Par ailleurs, pour tout nombre $x \in \mathbb{Z}$ on va utiliser la notation $\overline{x} = R(x)$. Il ne faut pas confondre avec la conjugaison complexe, mais puisque la proposition 2.24 nous dit que

$$\overline{x+y} = \overline{x} + \overline{y} \quad \text{et} \quad \overline{xy} = \overline{x}\,\overline{y}\,,$$

la notation se comporte comme prévu. De plus ceci est cohérent avec les « barres » que nous avons déjà, puisque $R(x) = x$ pour $0 \leq x < N$.

C'est la présence des barres qui, par convention, signifie que les calculs sont faits avec un modulo, qui en général est sous-entendu. Certains auteurs, pour éviter les ambiguïtés, préfèrent utiliser une notation plus complète, du genre

$$8 + 4 \equiv 0 \quad (12),$$

ou encore

$$8 + 4 \equiv 0 \mod 12,$$

là où, dans ce cours, nous écrivons

$$\overline{8} + \overline{4} = \overline{0},$$

le « 12 » étant sous-entendu. Mentionnons par ailleurs l'expression « x est congru à y modulo N » pour indiquer que $\overline{x} = \overline{y}$ dans $\mathbb{Z}/N\mathbb{Z}$.

voir les exercices 3151, 3152

Deuxième partie

Analyse

Chapitre 3

Suites

Définition 3.1 – Une *suite* de nombres réels est simplement une fonction $u\colon \mathbb{N} \to \mathbb{R}$. En général on écrit u_n au lieu de $u(n)$, et on écrit $(u_n)_{n\geq 0}$ pour désigner la suite elle-même.

Exemple 3.2 – La suite définie par $u_n = n^2$ commence par

$$0, 1, 4, 9, 16, 25, 36, \ldots$$

On emploie souvent une formule directe pour u_n en fonction de n, et dans ce cas on parle directement de « la suite $(n^2)_{n\geq 0}$ ».

On s'autorise aussi à parler de suites qui ne sont définies que pour des valeurs de n suffisamment grandes ; ainsi de la suite $(\frac{1}{n})_{n\geq 1}$ par exemple. On veillera à toujours indiquer le domaine de définition, ici l'ensemble des entiers ≥ 1. Évidemment l'étude de cette suite se ramène à celle de $(\frac{1}{n+1})_{n\geq 0}$: dans les deux cas il s'agit de comprendre la séquence de nombres

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$$

L'écriture $(\frac{1}{n})_{n\geq 1}$ est donc juste une notation commode.

Exemple 3.3 – Une autre façon commune de décrire une suite est d'utiliser une *relation de récurrence* : par exemple, on peut considérer la suite $(u_n)_{n\geq 0}$ définie par $u_0 = 1$ et $u_{n+1} = 2u_n$. Elle commence par

$$1, 2, 4, 8, 16, 32, 64, 128, \ldots$$

Dans cet exemple on voit tout de suite que $u_n = 2^n$. Pour le démontrer, sans surprise on procède par récurrence : on a bien $u_0 = 2^0 = 1$ et si $u_n = 2^n$ alors $u_{n+1} = 2 \cdot 2^n = 2^{n+1}$.

Les choses sont en général bien plus compliquées. Que l'on considère la suite $(v_n)_{n\geq 0}$ définie par à peu près n'importe quelle formule de récurrence choisie au hasard, disons $v_{n+1} = \cos(v_n) + (\sin(v_n))^3$, et l'on verra la difficulté qu'il peut y avoir à trouver une formule pour v_n.

Exemple 3.4 – On peut aussi utiliser une relation de récurrence qui fait intervenir plusieurs termes antérieurs. La célèbre *suite de Fibonacci* est définie par $u_0 = u_1 = 1$ et par $u_{n+2} = u_{n+1} + u_n$. Elle commence donc par

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

Bien plus tard nous verrons comment trouver une expression pour u_n dans le cas de cette suite (voir page 129).

Exemple 3.5 – Soit $(a_n)_{n\geq 0}$ une première suite. On peut alors définir $(u_n)_{n\geq 0}$ par

$$u_n = a_0 + a_1 + \cdots + a_n = \sum_{k=0}^{n} a_k \,.$$

On dit alors que $(u_n)_{n\geq 0}$ est la *série de terme général a_n*.

Par exemple lorsque $a_n = r^n$ pour un réel r, de sorte que $u_n = 1 + r + r^2 + \cdots + r^n$, on dit que (u_n) est une *série géométrique* (de raison r). On peut très facilement trouver une expression pour u_n en remarquant que

$$u_n(1-r) = (1 + r + \cdots + r^n) - (r + r^2 + \cdots r^{n+1}) = 1 - r^{n+1} \,.$$

Si $r \neq 1$, on en déduit

$$u_n = \sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r} \,.$$

C'est une formule qui sert tout le temps, elle est donc à savoir.

C'est à l'aide des suites que l'on va pouvoir traduire mathématiquement diverses notions de rapprochement : quantité qui s'approche « infiniment près » de 0, courbes qui se rapprochent « à l'infini », droites et cercles tangents à une courbe, et tant d'autres idées intuitives vont d'une façon ou d'une autre se ramener à des questions de suites.

La définition ci-dessous est au cœur de nombreux concepts dans ce livre. Il est donc normal qu'elle paraisse un peu difficile à saisir au début, et il faut prendre le temps de l'apprivoiser. Il s'agit de donner un sens à l'idée d'une suite qui se rapprocherait aussi près que l'on souhaite d'une valeur donnée. La formulation finale est due à Cauchy.

Définition 3.6 – Soit $(u_n)_{n\geq 0}$ une suite de nombres réels, et soit $\ell \in \mathbb{R}$. On dit que (u_n) *converge vers* ℓ, ou *admet* ℓ *pour limite*, lorsque la condition suivante est satisfaite : pour tout $\varepsilon > 0$ il doit y avoir un entier N_ε tel que $|u_n - \ell| < \varepsilon$ dès que $n \geq N_\varepsilon$.

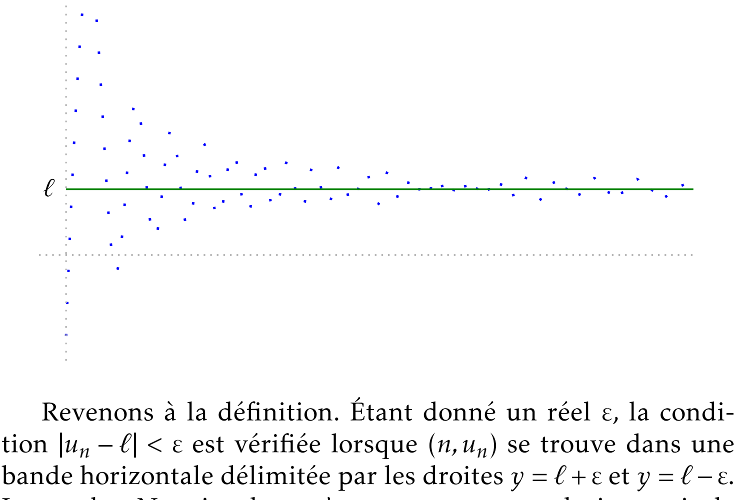
En d'autres termes, pour toute marge d'erreur ε donnée, la distance entre u_n et ℓ va devenir inférieure à ε pour peu que l'on prenne des indices suffisamment grands.

Dans ce cas on note

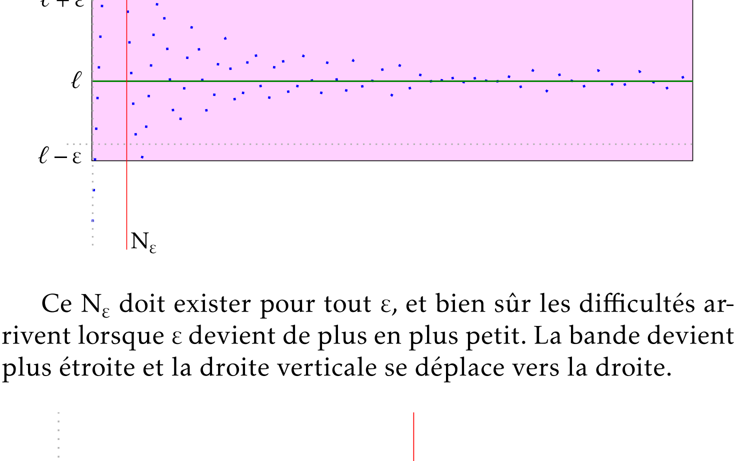
$$u_n \xrightarrow[n\to\infty]{} \ell \qquad \text{ou} \qquad \lim_{n\to\infty} u_n = \ell\,.$$

Essayons de comprendre cette définition graphiquement, pour commencer. Et d'abord, comment dessiner une suite ?

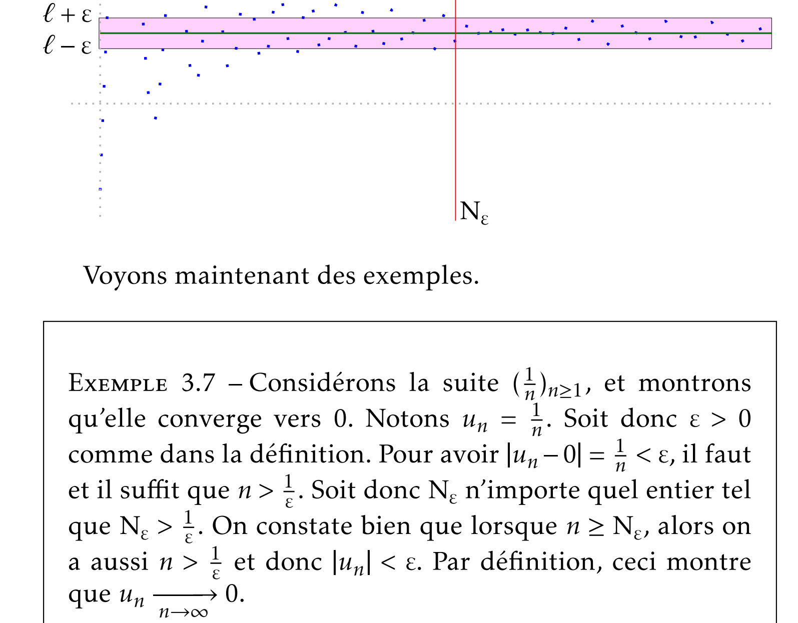
Nous allons procéder comme sur la figure ci-dessous. Les points représentés sont ceux de la forme (n, u_n), c'est-à-dire que le diagramme se lit de la gauche vers la droite à mesure que les indices augmentent. On a tracé les axes dans le plan $\mathbb{R}\times\mathbb{R}$, ainsi qu'une droite d'équation $y = \ell$ vers laquelle la suite semble s'accumuler. C'est une bonne façon de visualiser la convergence vers un nombre ℓ.



Revenons à la définition. Étant donné un réel ε, la condition $|u_n - \ell| < \varepsilon$ est vérifiée lorsque (n, u_n) se trouve dans une bande horizontale délimitée par les droites $y = \ell + \varepsilon$ et $y = \ell - \varepsilon$. Le nombre N_ε existe lorsqu'on peut tracer une droite verticale comme ci-dessous, d'équation $x = N_\varepsilon$, à la droite de laquelle tous les points (n, u_n) sans exception sont dans la bande.



Ce N_ε doit exister pour tout ε, et bien sûr les difficultés arrivent lorsque ε devient de plus en plus petit. La bande devient plus étroite et la droite verticale se déplace vers la droite.



Voyons maintenant des exemples.

Exemple 3.7 – Considérons la suite $(\frac{1}{n})_{n\geq 1}$, et montrons qu'elle converge vers 0. Notons $u_n = \frac{1}{n}$. Soit donc $\varepsilon > 0$ comme dans la définition. Pour avoir $|u_n - 0| = \frac{1}{n} < \varepsilon$, il faut et il suffit que $n > \frac{1}{\varepsilon}$. Soit donc N_ε n'importe quel entier tel que $N_\varepsilon > \frac{1}{\varepsilon}$. On constate bien que lorsque $n \geq N_\varepsilon$, alors on a aussi $n > \frac{1}{\varepsilon}$ et donc $|u_n| < \varepsilon$. Par définition, ceci montre que $u_n \xrightarrow[n\to\infty]{} 0$.

(Voir l'encadré « $\mathbb{R}$ est archimédien » pour quelques commentaires sur cet exemple.)

$\mathbb{R}$ est archimédien

Dans l'exemple 3.7, on utilise le fait suivant : étant donné un réel x, il existe un entier N tel que $N > x$ (dans l'exemple on avait $x = \frac{1}{\varepsilon}$ et $N = N_\varepsilon$).

D'abord, pourquoi est-ce vrai ? Le (4) du théorème 2.3 affirme qu'il existe un rationnel $\frac{p}{q}$ tel que $x < \frac{p}{q}$; il suffit alors de prendre $N = p$.

C'est Archimède qui le premier avait énoncé : « Pour deux grandeurs inégales, il existe toujours un multiple entier de la plus petite, supérieur à la plus grande. » En clair, étant donnés a et b rationnels ou réels (mais Archimède pensait seulement aux rationnels), tels que $0 < a < b$, alors il existe un entier n tel que $na > b$. (Ce qui revient à prendre $n > \frac{b}{a}$.) En particulier, la distance Terre-Lune peut être couverte par des allumettes mises bout à bout. Ou même côte à côte. C'est pour ces raisons historiques que l'on dit que $\mathbb{R}$ est *archimédien* quand on veut faire référence à cette propriété.

À titre d'exercice, vous montrerez que l'énoncé selon lequel $\mathbb{R}$ est archimédien équivaut à dire que pour tout $a, b \in \mathbb{R}$ avec $a < b$, il existe une fraction $f \in \mathbb{Q}$ telle que $a < f < b$.

Exemple 3.8 – Maintenant voyons $(\alpha^n)_{n\geq 0}$ pour un réel $0 \leq \alpha < 1$. Montrons que la suite tend encore vers 0. Puisque $|\alpha^n - 0| = |\alpha^n| = \alpha^n$, il s'agit de majorer les termes de la suite par quelque chose de facile à comprendre.

Prenons un nombre rationnel $\frac{p}{q}$ tel que $\alpha < \frac{p}{q} < 1$ (voir (3) du théorème 2.3). On a $q > p$ et comme il s'agit d'entiers, on est même sûr que $q \geq p + 1$; ainsi

$$\alpha^n < \frac{p^n}{q^n} < \frac{p^n}{(p+1)^n}\,.$$

Remarquons que

$$\begin{aligned} (p+1)^n &= (p+1)(p+1)\cdots(p+1) \\ &= p^n + np^{n-1} + \text{(termes } > 0) \\ &> p^n + np^{n-1}\,. \end{aligned}$$

On a donc $\frac{(p+1)^n}{p^n} > 1 + \frac{n}{p} > \frac{n}{p}$, et par suite

$$\alpha^n < \frac{p}{n}\,.$$

La suite $(\frac{p}{n})_{n\geq 1}$ tend vers 0, exactement comme la suite $(\frac{1}{n})_{n\geq 1}$ (la constante p ne change rien à l'affaire). Donc étant donné $\varepsilon > 0$, il existe bien N_ε tel que $\frac{p}{n} < \varepsilon$ pour $n \geq N_\varepsilon$; pour ces mêmes valeurs de n, on a donc aussi $\alpha^n < \varepsilon$, et finalement $\alpha^n \xrightarrow[n\to\infty]{} 0$.

Exemple 3.9 – On note $n! = n(n-1)(n-2)\cdots 3\times 2\times 1$, et on appelle cette quantité « *factorielle n* ». C'est le produit de n termes qui sont tous ≥ 2 sauf le dernier ; on a donc $n! \geq 2^{n-1}$.

Si l'on étudie la suite $(\frac{1}{n!})_{n\geq 1}$, il suffit de remarquer que

$$0 < \frac{1}{n!} \leq \frac{1}{2^{n-1}} = \left(\frac{1}{2}\right)^{n-1} \qquad \text{et donc} \qquad \frac{1}{n!} \xrightarrow[n\to\infty]{} 0\,.$$

En effet l'exemple précédent, dans le cas $\alpha = \frac{1}{2}$, montre que $\frac{1}{2^n}$ converge vers 0 ; or si on a une suite u_n qui tend vers 0, et si $0 \leq v_n \leq u_n$, alors v_n tend aussi vers 0. Vérifiez ceci à partir de la définition, puis habituez-vous à faire ce genre de petit raisonnement rapidement.

voir les exercices 519, 505, 524

Les exemples précédents sont presque les seuls pour lesquels vous verrez un ε cette année. La quasi-totalité des suites que l'on va rencontrer vont être des combinaisons de ces limites, pour lesquelles on va appliquer le résultat suivant.

Proposition 3.10 – *Soient $(u_n)_{n\geq 0}$ et $(v_n)_{n\geq 0}$ deux suites. On suppose que $u_n \xrightarrow[n\to\infty]{} \ell$ et que $v_n \xrightarrow[n\to\infty]{} \ell'$. Alors on a :*

1. *(somme) $u_n + v_n \xrightarrow[n\to\infty]{} \ell + \ell'$;*
2. *(produit) $u_n v_n \xrightarrow[n\to\infty]{} \ell\ell'$;*
3. *(inverse) $\dfrac{1}{u_n} \xrightarrow[n\to\infty]{} \dfrac{1}{\ell}$ lorsque $\ell \neq 0$.*

Avant de donner la démonstration, montrons un petit résultat intermédiaire :

Lemme 3.11 – *Toute suite qui converge est bornée. En d'autres termes, si $(u_n)_{n\geq 0}$ admet une limite, alors il existe un nombre C tel que pour tout entier n, on a $|u_n| \leq C$.*

De plus, si la limite est > 0, alors il existe une constante $\rho > 0$ telle que $u_n > \rho$ pour tous les n assez grands.

Démonstration. Soit ℓ la limite. On écrit

$$|u_n| = |(u_n - \ell) + \ell| \leq |u_n - \ell| + |\ell|.$$

Soit $\varepsilon > 0$, par exemple $\varepsilon = 1$. Il existe un entier N_ε tel que pour tous les $n \geq N_\varepsilon$, on a $|u_n - \ell| < \varepsilon$. Il suffit alors de prendre C plus grand que $\varepsilon + |\ell|$, et plus grand que tous les nombres $|u_n|$ pour $n < N_\varepsilon$ (qui sont en nombre fini).

Si maintenant $\ell > 0$, posons $\varepsilon = \frac{\ell}{2}$. Pour $n \geq N_\varepsilon$ on a $|u_n - \ell| < \varepsilon$, ce qu'on va réécrire

$$\frac{\ell}{2} = \ell - \varepsilon < u_n < \ell + \varepsilon.$$

On peut donc prendre $\rho = \varepsilon = \frac{\ell}{2}$. □

Démonstration de la proposition. Commençons par la formule pour le produit. Une astuce que vous reverrez souvent est d'écrire $u_n v_n - \ell\ell' = u_n(v_n - \ell') + \ell'(u_n - \ell)$. Par suite

$$|u_n v_n - \ell\ell'| \leq |u_n| \cdot |v_n - \ell'| + |\ell'| \cdot |u_n - \ell|. \qquad (*)$$

Soit donc $\varepsilon > 0$. Par hypothèse il existe N_1 tel que

$$|u_n - \ell| < \varepsilon \qquad (1)$$

dès que $n \geq N_1$; de même on a un N_2 tel que

$$|v_n - \ell'| < \varepsilon \qquad (2)$$

dès que $n \geq N_2$.

Si nous prenons N_ε n'importe quel nombre à la fois plus grand que N_1 et plus grand que N_2, alors on a à la fois (1) et (2) lorsque $n \geq N_\varepsilon$. Prenons une constante C comme dans le lemme. Alors en reportant ces inégalités dans (*), on aboutit à

$$|u_n v_n - \ell\ell'| \leq (C + |\ell'|)\varepsilon \quad \text{pour} \quad n \geq N_\varepsilon. \qquad (**)$$

Avec l'habitude, vous vous rendrez compte que ce genre d'argument est suffisant, et que l'on a essentiellement déjà montré que $u_n v_n$ a pour limite $\ell\ell'$. Pourquoi ?

Simplement parce que pour tout $\varepsilon > 0$, on vient de montrer que l'on sait trouver N_ε tel que l'inégalité (**) est valable. En particulier on peut faire ce travail pour $\tilde\varepsilon = \frac{\varepsilon}{C+|\ell'|}$; donc il existe $N_{\tilde\varepsilon}$ tel que pour les entiers $n \geq N_{\tilde\varepsilon}$ on a effectivement

$$|u_n v_n - \ell\ell'| \leq (C + |\ell'|)\tilde\varepsilon = \varepsilon.$$

Vous montrerez la formule pour la somme sur le même modèle, en plus facile.

Pour l'inverse on écrit

$$\left|\frac{1}{u_n} - \frac{1}{\ell}\right| = \left|\frac{\ell - u_n}{u_n \ell}\right| < \frac{1}{\rho|\ell|}|u_n - \ell|,$$

où ρ est comme dans le lemme (appliqué à la suite $|u_n|$ qui converge vers $|\ell| > 0$ (vérifiez-le)). Ainsi pour tout $\varepsilon > 0$, on trouve un N_ε tel que pour $n \geq N_\varepsilon$ on a

$$\left|\frac{1}{u_n} - \frac{1}{\ell}\right| < \frac{\varepsilon}{\rho|\ell|}.$$

Là encore, c'est suffisant pour affirmer que $\frac{1}{u_n}$ converge vers $\frac{1}{\ell}$. □

Exemple 3.12 – Voyons comment cette proposition nous simplifie la vie. Admettons que l'on souhaite connaître la limite de

$$u_n = \frac{4n^2 + 1}{5n^2 - n + 2}.$$

On commence par diviser par n^2 au numérateur comme au dénominateur :

$$u_n = \frac{4 + \frac{1}{n^2}}{5 - \frac{1}{n} + \frac{2}{n^2}}.$$

On a vu dans l'exemple 3.7 que $\frac{1}{n} \to 0$. Grâce à la formule pour le produit, on sait désormais que $\frac{1}{n^2} = (\frac{1}{n})(\frac{1}{n})$ converge également vers $0 \times 0 = 0$. De même $\frac{2}{n^2} \to 2 \times 0 = 0$ et $-\frac{1}{n} \to (-1) \times 0 = 0$, encore par la formule pour le produit.

Maintenant, la formule pour la somme nous dit que $4 + \frac{1}{n^2} \to 4$ et que $5 - \frac{1}{n} + \frac{2}{n^2} \to 5 \neq 0$. On peut donc utiliser la formule pour l'inverse qui donne finalement

$$u_n \xrightarrow[n\to\infty]{} \frac{4}{5}.$$

voir les exercices 507, 563, 568

DÉFINITION 3.13 – On dit que la suite $(u_n)_{n\geq 0}$ est *croissante* lorsque $u_n \leq u_{n+1}$ pour tout n ; on dit qu'elle est *décroissante* lorsque $u_{n+1} \leq u_n$.

On dit que (u_n) est *majorée* lorsqu'il existe $M \in \mathbb{R}$ tel que $u_n \leq M$, pour tout n ; on dit qu'elle est *minorée* lorsqu'il existe m tel que $m \leq u_n$.

THÉORÈME 3.14 – *Toute suite croissante et majorée est convergente.*

Toute suite décroissante et minorée est convergente.

La démonstration va faire appel à la notion de borne supérieure, introduite dans le chapitre « Nombres ». Pour ceux qui n'ont pas lu ce chapitre, précisons que le théorème 3.14 (que vous allez devoir admettre) est essentiellement équivalent à celui qui affirme que les bornes supérieures existent dans $\mathbb{R}$. En d'autres termes, presque à chaque fois qu'une démonstration reposera sur une utilisation de sup, on pourra la réécrire différemment avec des suites monotones.

Démonstration. Soit $(u_n)_{n\geq 0}$ une suite croissante et majorée. On pose

$$\ell = \sup\{u_n \mid n \in \mathbb{N}\},$$

ce qui a un sens puisque cet ensemble est majoré par hypothèse.

Soit $\varepsilon > 0$ et considérons $\ell' = \ell - \varepsilon < \ell$. Alors ℓ' ne peut pas être un majorant de l'ensemble ci-dessus, puisque ℓ est le plus petit. Ce qui revient à dire qu'il y a au moins un élément de l'ensemble, disons u_N, tel que $u_N > \ell'$.

Or la suite est croissante, donc $u_n \geq u_N > \ell'$ pour tous les $n \geq N$. On a donc $\ell - \varepsilon < u_n \leq \ell$ pour ces valeurs de n, et finalement $|u_n - \ell| < \varepsilon$. On peut donc prendre $N_\varepsilon = N$ et la suite converge vers le sup de ses valeurs.

Si (u_n) est décroissante et minorée, on applique le résultat ci-dessus à $(-u_n)$, qui est croissante et majorée. □

EXEMPLE 3.15 – Voici une deuxième façon, plus facile, de monter que $\alpha^n \to 0$ lorsque $0 < \alpha < 1$. Posons $u_n = \alpha^n$. Alors $\frac{u_{n+1}}{u_n} = \alpha < 1$, donc $u_{n+1} < u_n$: la suite est décroissante. Tous les termes sont > 0, donc elle est minorée. Par le théorème, (u_n) admet une certaine limite ℓ. Montrons que $\ell = 0$.

Soit $v_n = \alpha^{n+1}$. D'un côté nous avons $v_n = \alpha\,\alpha^n = \alpha u_n$. Par la formule pour les produits de suites, on en déduit $v_n \xrightarrow[n\to\infty]{} \alpha\ell$.

D'un autre côté, nous avons $v_n = u_{n+1}$. Il est donc clair que v_n a la même limite que u_n puisque c'est la même suite avec simplement les termes décalés d'un cran. Donc $v_n \xrightarrow[n\to\infty]{} \ell$.

On doit donc avoir $\ell = \alpha\ell$. Si on avait $\ell \neq 0$, on en déduirait $\alpha = 1$, ce qui contredit les hypothèses. Donc $\ell = 0$.

voir les exercices 569, 572, 574

Définition 3.16 – On écrit

$$u_n \xrightarrow[n\to\infty]{} +\infty \quad \text{ou} \quad \lim_{n\to\infty} u_n = +\infty,$$

lorsque la condition suivante est remplie. Pour tout $M > 0$, il doit exister un entier $N = N_M$ tel que $u_n > M$ pour tous les entiers $n \geq N$.

De même on dit que $u_n \to -\infty$ lorsque pour tout $m < 0$, il existe un entier $N = N_m$ tel que $u_n < m$ dès que $n \geq N$.

En d'autres termes, quand $u_n \to +\infty$ les termes de la suite deviennent arbitrairement grands lorsque les indices sont suffisamment grands.

Les exemples vont provenir du résultat suivant. C'est une variante de la proposition 3.10, mais les choses ne marchent pas aussi bien.

Proposition 3.17 – *Soient $(u_n)_{n\geq 0}$ et $(v_n)_{n\geq 0}$ deux suites. On suppose que $u_n \xrightarrow[n\to\infty]{} +\infty$ et que $v_n \xrightarrow[n\to\infty]{} \ell$, où l'on peut avoir aussi bien $\ell \in \mathbb{R}$ que $\ell = \pm\infty$. Alors :*

1. *(somme) si $\ell \neq -\infty$ alors $u_n + v_n \to +\infty$;*
2. *(produit)*
   (a) *si $\ell > 0$ alors $u_n v_n \xrightarrow[n\to\infty]{} +\infty$, et*
   (b) *si $\ell < 0$ alors $u_n v_n \xrightarrow[n\to\infty]{} -\infty$;*
3. *(inverse)*
   (a) $\dfrac{1}{u_n} \xrightarrow[n\to\infty]{} 0$;
   (b) *si $\ell = 0$, et si $\forall n$ on a $v_n > 0$, alors $\dfrac{1}{v_n} \xrightarrow[n\to\infty]{} +\infty$;*
   (c) *si $\ell = 0$, et si $\forall n$ on a $v_n < 0$, alors $\dfrac{1}{v_n} \xrightarrow[n\to\infty]{} -\infty$.*

Vous montrerez cette proposition à titre d'exercice. Avant de passer aux exemples, notons que tous les cas ne sont pas couverts par ce dernier énoncé : que dire de $u_n + v_n$ lorsque $\ell = -\infty$? Que dire de $u_n v_n$ lorsque $\ell = 0$? Réponse : rien en général. On dit que ce sont les « formes indéterminées ». Toutes les situations sont envisageables.

Exemple 3.18 – Puisque $\frac{1}{n^k} \xrightarrow[n\to\infty]{} 0$ pour tout entier k , le 3(b) de la proposition nous dit que $n^k \xrightarrow[n\to\infty]{} +\infty$. On peut évidemment vérifier très facilement ceci à partir de la définition.

De la même manière, $\beta^n \to +\infty$ si $\beta > 1$: en effet posons $\alpha = \frac{1}{\beta} < 1$, alors $\alpha^n \to 0$, et on applique encore le 3(b). Toujours pour les mêmes raisons, on a $n! \to +\infty$.

Enfin, nous avons vu un exemple de forme indéterminée dans l'exemple 3.12, en l'occurence une « forme $\frac{+\infty}{+\infty}$ ». Nous avions montré dans ce cas précis qu'il y avait bien une limite, à savoir $4/5$.

Il y a certaines formes indéterminées que l'on peut résoudre, et il est utile d'en mémoriser quelques-unes au fur et à mesure qu'on les rencontre. En voici une première, « l'exponentielle l'emporte sur les puissances de n » :

Lemme 3.19 – *Soit α un réel tel que $0 \leq \alpha < 1$, et k un entier. Alors*

$$n^k \alpha^n \xrightarrow[n\to\infty]{} 0.$$

Démonstration. C'est une forme indéterminée du type « $+\infty \times 0$ ». Soit $u_n = n^k \alpha^n$. On fait l'estimation suivante :

$$\frac{u_{n+1}}{u_n} = (1 + \frac{1}{n})^k \alpha \xrightarrow[n\to\infty]{} \alpha.$$

Prenons $\varepsilon = \frac{1-\alpha}{2} > 0$; alors pour tous les $n \geq N$, pour un certain $N = N_\varepsilon$, on a

$$\alpha - \varepsilon < \frac{u_{n+1}}{u_n} < \alpha + \varepsilon = \frac{\alpha + 1}{2} < 1.$$

En posant $\rho = \frac{\alpha+1}{2}$, on voit d'abord que $u_{N+1} < \rho u_N$; puis $u_{N+2} < \rho u_{N+1} < \rho^2 u_N$; et une récurrence nous mène immédiatement à

$$u_{N+n} < \rho^n u_N,$$

pour tous les $n \geq 0$. Comme $\rho^n u_N \xrightarrow[n\to\infty]{} 0$ (le terme u_N n'est qu'une constante !), on en déduit bien que u_n converge vers 0. □

Dans le même genre, « factorielle l'emporte sur l'exponentielle » :

Lemme 3.20 – *Soit α un réel tel que $0 \leq \alpha < 1$. Alors*

$$\frac{1}{\alpha^n n!} \xrightarrow[n\to\infty]{} 0.$$

Démonstration. Posons $u_n = \frac{1}{\alpha^n n!}$. Alors

$$\frac{u_{n+1}}{u_n} = \frac{1}{\alpha(n+1)} \xrightarrow[n\to\infty]{} 0,$$

donc $u_{n+1} < u_n$ pour n suffisamment grand. La suite (u_n), qui finit par être décroissante et minorée, converge vers $\ell \geq 0$. Or si $\ell > 0$, on aurait

$$\frac{u_{n+1}}{u_n} \xrightarrow[n\to\infty]{} \frac{\ell}{\ell} = 1,$$

une contradiction. □

– Fin de la première lecture –

Partant d'une première suite $(a_n)_{n\geq 0}$, on peut considérer la « série de terme général a_n », c'est-à-dire la suite

$$u_n = a_0 + a_1 + a_2 + \cdots + a_n = \sum_{k=0}^{n} a_k .$$

(Cf. l'exemple 3.5.)

Le théorème suivant peut paraître surprenant : il dit que pour montrer la convergence de u_n, il suffit de montrer la convergence de la série de terme général $|a_n|$:

Théorème 3.21 – *Soit $(a_n)_{n\geq 0}$ une suite. Si la limite*

$$\lim_{n\to\infty} \sum_{k=0}^{n} |a_k|$$

existe, alors la limite

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k$$

existe également. On dit alors que la série de terme général a_n converge absolument.

Démonstration. On pose

$$S_n = \sum_{k=0}^{n} |a_k| \quad \text{et} \quad u_n = \sum_{k=0}^{n} a_k .$$

Par hypothèse $S_n \to \ell$. Soit $\varepsilon > 0$, alors pour tous les n suffisamment grands on aura $|S_n - \ell| < \frac{\varepsilon}{2}$, et donc

$$|S_n - S_m| = |(S_n - \ell) + (\ell - S_m)| \leq |S_n - \ell| + |S_m c - \ell| < \varepsilon$$

lorsque n et m sont tous les deux supérieurs à un certain N. On en déduit pour la suite (u_n) que

$$|u_n - u_m| = \left| \sum_{k=m+1}^{n} a_k \right| \leq \sum_{k=m+1}^{n} |a_k| = S_n - S_m < \varepsilon,$$

lorsque $n \geq m \geq$ N. On dit souvent que (u_n) est une suite *de Cauchy* pour exprimer cette propriété (à savoir que $|u_n - u_m| < \varepsilon$ pour n et m suffisamment grands). La fin de la démonstration va établir qu'une suite de Cauchy de nombres réels converge toujours.

En effet, soient

$$\alpha_n = \inf\{u_k \mid k \geq n\} \quad \text{et} \quad \beta_n = \sup\{u_k \mid k \geq n\},$$

de sorte que $\alpha_n \leq u_n \leq \beta_n$. Par construction la suite α_n est croissante et majorée ; donc elle admet une limite ℓ_1 d'après le théorème 3.14. De même β_n converge vers une limite ℓ_2 parce qu'elle est décroissante et minorée.

Mais nous avons montré que pour tout $\varepsilon > 0$, il existe un rang N au-delà duquel $|u_n - u_N| < \varepsilon$; pour ces valeurs de n, le nombre u_n est dans l'intervalle $]u_N - \varepsilon; u_N + \varepsilon[$ de longueur 2ε, et on en déduit que $\beta_n - \alpha_n \leq 2\varepsilon$ pour $n \geq$ N. Ceci montre que $\beta_n - \alpha_n \to 0$ et donc que $\ell_1 = \ell_2$. Finalement l'encadrement $\alpha_n \leq u_n \leq \beta_n$ garantit que u_n converge également vers cette limite. □

Notez bien que la démonstration ne dit pas du tout comment *calculer* la limite de $\sum a_k$, même si on connaît la limite de $\sum |a_k|$. Notez également que ce théorème serait faux si on travaillait sur $\mathbb{Q}$; d'ailleurs la démonstration utilise des bornes supérieures et inférieures, qui sont propres à $\mathbb{R}$.

Dorénavant, nous utiliserons la notation suivante, plus suggestive, pour les limites de sommes. On écrit :

$$\sum_{k=0}^{+\infty} a_k = \lim_{n\to\infty} \sum_{k=0}^{n} a_k ,$$

lorsque cette limite existe.

Exemple 3.22 (L'exponentielle) – Soit $x \in \mathbb{R}$ fixé. Posons

$$u_n = \sum_{k=0}^{n} \frac{x^k}{k!} .$$

Montrons que cette suite converge absolument. Nous devons donc montrer que

$$S_n = \sum_{k=0}^{n} \frac{|x|^k}{k!}$$

admet une limite. La suite (S_n) est croissante, donc d'après le théorème 3.14 il suffit de montrer qu'elle est majorée.

Pour k suffisamment grand, disons $k \geq$ K, on a $|x| < k$; on peut même choisir $\alpha < 1$ tel que $\frac{|x|}{k} < \alpha$ pour $k \geq$ K. On a alors

$$\frac{|x|^{K+1}}{(K+1)!} = \frac{|x|}{K+1} \cdot \frac{|x|^K}{K!} < \alpha \frac{|x|^K}{K!} .$$

De même on a

$$\frac{|x|^{K+2}}{(K+2)!} = \frac{|x|}{K+2} \cdot \frac{|x|^{K+1}}{(K+1)!} < \alpha \frac{|x|^{K+1}}{(K+1)!} < \alpha^2 \frac{|x|^K}{K!} .$$

Par récurrence on obtient

$$\frac{|x|^{K+k}}{(K+k)!} < \alpha^k \frac{|x|^K}{K!} .$$

Ceci va nous suffire, puisqu'en posant $C = S_{K-1}$ on peut écrire pour $n \geq$ K :

$$\begin{aligned} S_n &= C + \sum_{k=0}^{n-K} \frac{|x|^{K+k}}{(K+k)!} \\ &\leq C + \frac{|x|^K}{K!}(1 + \alpha + \alpha^2 + \cdots + \alpha^{n-K}) \\ &= C + \frac{|x|^K}{K!} \frac{1 - \alpha^{n+1-K}}{1 - \alpha} \\ &\leq C + \frac{|x|^K}{K!} \frac{1}{1 - \alpha} . \end{aligned}$$

La suite (S_n) est donc bien majorée en plus d'être croissante, elle est donc convergente ; par suite (u_n) est absolument convergente, et donc elle-même convergente d'après le théorème. Sa limite, qui dépend de x, est notée $\exp(x)$ ou e^x, et appelée *l'exponentielle* du nombre x. En clair

$$e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!} .$$

Cette définition de l'exponentielle coïncide, heureusement, avec les définitions qui vous sont esquissées au lycée. Nous montrerons ça en temps voulu.

voir l'exercice 570

Exemple 3.23 – Certaines séries convergent, mais sans converger absolument : il faut donc faire attention. Par exemple, nous pourrons montrer plus loin que pour $a_k = \frac{(-1)^k}{k+1}$, la série converge et on a même

$$\sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} + \cdots = \ln(2).$$

Par contre en prenant les valeurs absolues, nous montrerons que l'on a en fait

$$\sum_{k=0}^{+\infty} \frac{1}{k+1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots = +\infty,$$

c'est-à-dire qu'il n'y a pas de limite finie.

Lorsque l'on se donne pour chaque entier n un nombre complexe $z_n = a_n + ib_n$, et lorsque $a_n \to \ell_1$ et $b_n \to \ell_2$, on dit que $(z_n)_{n\geq 0}$ converge vers $\ell = \ell_1 + i\ell_2$, et on note $z_n \xrightarrow[n\to\infty]{} \ell$. Par exemple

$$\frac{3n^2}{2^n} + \frac{2n+5}{7n-12}i \xrightarrow[n\to\infty]{} \frac{2i}{7}.$$

Cette définition a le mérite d'être simple. Cependant on peut donner une définition plus directe, sans référence aux suites réelles, en remplaçant simplement les valeurs absolues par les modules ; en clair :

Proposition 3.24 – *La suite $(z_n)_{n\geq 0}$ converge vers ℓ exactement lorsque la condition suivante est remplie. Pour chaque réel $\varepsilon > 0$, il doit y avoir un entier N_ε tel que $|z_n - \ell| < \varepsilon$ dès que $n \geq N_\varepsilon$.*

Ici $|z_n - \ell|$ est le module du nombre complexe $z_n - \ell$. À part ça, la condition est la même que pour les suites réelles.

Démonstration. Écrivons $z_n = a_n + ib_n$. Pour commencer, supposons que $a_n \to \ell_1$ et $b_n \to \ell_2$. Étant donné $\varepsilon > 0$, on trouve N_1 tel que $|a_n - \ell_1| < \varepsilon$ pour $n \geq N_1$, et de même pour tous les n plus grands qu'un certain entier N_2 on a $|b_n - \ell_2| < \varepsilon$. Lorsque n est à la fois plus grand que N_1 et que N_2, on a

$$|z_n - (\ell_1 + i\ell_2)| = \sqrt{(a_n - \ell_1)^2 + (b_n - \ell_2)^2} \leq \sqrt{\varepsilon^2 + \varepsilon^2} \leq \sqrt{2}\varepsilon.$$

Ceci montre bien (en recommençant avec $\tilde{\varepsilon} = \frac{\varepsilon}{\sqrt{2}}$) que la condition donnée est remplie pour $\ell = \ell_1 + i\ell_2$.

Réciproquement, supposons cette condition remplie pour un certain $\ell = \ell_1 + i\ell_2$, et étudions la convergence des suites (a_n) et (b_n). Soit $\varepsilon > 0$. En notant simplement que

$$|a_n - \ell_1| \leq |z_n - \ell|,$$

on constate que si $|z_n - \ell| < \varepsilon$, alors $|a_n - \ell_1| < \varepsilon$, et ceci établit que $a_n \to \ell_1$, clairement. De même (b_n) converge vers ℓ_2. □

La convergence absolue fonctionne encore avec les complexes :

Théorème 3.25 – *Soit $(z_n)_{n\geq 0}$ une suite de nombres complexes. Si la limite*

$$\lim_{n\to\infty} \sum_{k=0}^{n} |z_k|$$

existe, alors la limite

$$\lim_{n\to\infty} \sum_{k=0}^{n} z_k$$

existe également. On dit que la série converge absolument.

Démonstration. Écrivons $z_n = a_n + ib_n$. On a $|a_n| \leq |z_n|$ et donc

$$S_n = \sum_{k=0}^{n} |a_k| \leq \sum_{k=0}^{n} |z_k| \leq \sum_{k=0}^{+\infty} |z_k|.$$

La suite S_n est donc majorée. Elle est visiblement croissante, donc elle converge. D'après le théorème 3.21, on en déduit l'existence de

$$\sum_{k=0}^{+\infty} a_k\,;$$

de même on montre l'existence de

$$\sum_{k=0}^{+\infty} b_k.$$

Mais bien sûr on a

$$\mathfrak{Re}\left(\sum_{k=0}^{n} z_k\right) = \sum_{k=0}^{n} a_k, \quad \mathfrak{Im}\left(\sum_{k=0}^{n} z_k\right) = \sum_{k=0}^{n} b_k.$$

Puisque ses parties réelles et imaginaires convergent, c'est bien que la somme elle-même converge. □

Exemple 3.26 (L'exponentielle complexe) – Soit $z \in \mathbb{C}$ fixé. On pose

$$u_n = \sum_{k=0}^{n} \frac{z^k}{k!}.$$

Soit $x = |z|$; c'est un nombre réel ≥ 0, et on a montré dans l'exemple 3.22 l'existence de

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!}.$$

C'est donc que (u_n) converge absolument. D'après le théorème, elle converge. La limite dépend de z, on la note $\exp(z)$ ou e^z. En clair

$$e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!}.$$

Lorsque $z = i\theta$ avec $\theta \in \mathbb{R}$, la notation $e^{i\theta}$ coïncide avec celle que vous connaissiez au lycée, comme nous le montrerons plus loin.

Sur cet exemple on peut apprécier le secours qui nous est apporté par le théorème sur la convergence absolue : étudier les parties réelle et imaginaire de (u_n) directement serait bien difficile.

Pour travailler directement avec les complexes sans passer par les parties réelle et imaginaire, il nous manque encore un ingrédient : c'est la très utile inégalité triangulaire, que nous avons vue dans le cas de $\mathbb{R}$ dans le lemme 2.4. Elle reste vraie sur $\mathbb{C}$:

Lemme 3.27 – *Si a et b sont des nombres complexes, on a*

$$|a + b| \leq |a| + |b|,$$

et

$$||a| - |b|| \leq |a - b|.$$

Nous allons donner une démonstration très générale dans le paragraphe suivant (voir le lemme 3.31).

L'ensemble $\mathbb{C}$ des nombres complexes peut être identifié avec l'ensemble $\mathbb{R}\times\mathbb{R}$, que l'on va noter $\mathbb{R}^2$, en voyant $a+ib$ comme la paire (a,b). De même, on peut considérer l'ensemble $\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ des triplets (a,b,c) de nombres réels ; on va noter cet ensemble $\mathbb{R}^3$. L'ensemble $\mathbb{R}^4$ est composé des quadriplets (a,b,c,d).

Rien ne nous enpêche de continuer : étant donné un entier r, l'ensemble $\mathbb{R}^r$ est constitué des « r-uplets » $(a,b,c,d,\ldots)$ (séquence de r nombres réels). Ces éléments sont appelés *vecteurs*.

DÉFINITION 3.28 – La *norme* (ou norme euclidienne) d'un vecteur est :

$$\|(a,b,c,d,\ldots)\| = \sqrt{a^2+b^2+c^2+d^2+\cdots}.$$

En particulier, si $z=a+ib$, alors $|z| = \|(a,b)\|$. La norme est donc une généralisation du module.

Une suite de vecteurs est une fonction $\mathbb{N}\to\mathbb{R}^r$, c'est-à-dire que pour tout entier n on se donne un vecteur

$$u_n = (a_n, b_n, c_n, d_n, \ldots) \in \mathbb{R}^r.$$

Exactement comme dans le cas des complexes, on a :

PROPOSITION 3.29 – *Soit $u_n = (a_n, b_n, c_n, d_n, \ldots)$ une suite de vecteurs de $\mathbb{R}^r$. Les deux énoncés ci-dessous sont équivalents :*

1. *Chacune des suites $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$, $(c_n)_{n\geq 0}$, ..., converge respectivement vers $\ell_1, \ell_2, \ell_3, \ldots$*
2. *Soit $\ell = (\ell_1, \ell_2, \ldots, \ell_r)$. Pour chaque réel $\varepsilon > 0$, il existe un entier N_ε tel que pour $n \geq N_\varepsilon$ on a $\|u_n - \ell\| < \varepsilon$.*

THÉORÈME 3.30 – *Soit $(a_n)_{n\geq 0}$ une suite de vecteurs dans $\mathbb{R}^r$. Si la limite*

$$\lim_{n\to\infty} \sum_{k=0}^{n} \|a_k\|$$

existe (dans $\mathbb{R}$), alors la limite

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k$$

existe également (et c'est un vecteur de $\mathbb{R}^r$). On dit que la série converge absolument.

Les démonstrations sont les mêmes que dans le cas des complexes, et sont laissées en exercice. Nous avons également :

LEMME 3.31 – *Si a et b sont des vecteurs de $\mathbb{R}^r$, alors*

$$\|a+b\| \leq \|a\| + \|b\|,$$

et

$$\big|\, \|a\| - \|b\| \,\big| \leq \|a-b\|.$$

Montrons-le (ceci va établir le lemme 3.27 du même coup). Commençons par une inégalité célèbre :

LEMME 3.32 (INÉGALITÉ DE CAUCHY-SCHWARZ) – *Étant donnés des nombres réels $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_r$, on a :*

$$\left|\sum_{i=1}^{r} x_i y_i\right|^2 \leq \left(\sum_{i=1}^{r} x_i^2\right)\left(\sum_{i=1}^{r} y_i^2\right).$$

Supposons de plus que les nombres $y_1, y_2, \ldots, y_r$ ne sont pas tous nuls ; alors cette inégalité est une égalité exactement lorsqu'il existe un réel t tel que $x_i + ty_i = 0$ pour tous les indices i à la fois.

Démonstration. Si tous les y_i sont nuls, les deux membres de l'inégalité sont nuls, et l'inégalité est donc satisfaite. On suppose maintenant qu'il ne sont pas tous nuls.

Pour $t\in\mathbb{R}$, considérons

$$P(t) = \sum_{i=0}^{r} (x_i + ty_i)^2 = At^2 + Bt + C,$$

avec

$$A = \sum_{i=0}^{r} y_i^2, \quad B = 2\sum_{i=0}^{r} x_i y_i, \quad C = \sum_{i=0}^{r} x_i^2.$$

Faisons quelques observation sur $P(t)$. D'abord, puisque $P(t)$ est une somme de carrés, on a $P(t)\geq 0$; de plus $P(t)=0$ exactement lorsque $x_i + ty_i = 0$ pour tous les indices i à la fois. Par hypothèse il y a un indice i_0 tel que $y_{i_0}\neq 0$, donc une seule valeur de t au maximum peut convenir, à savoir $t = -\frac{x_{i_0}}{y_{i_0}}$.

Concluons. Ou bien le polynôme $P(t)$ a une racine réelle et une seule, et donc son discriminant $B^2 - 4AC = 0$; ou bien il n'a pas de racine réelle du tout, et donc son discriminant $B^2 - 4AC < 0$. Étant données les valeurs de A, B et C, on a exactement ce que dit le lemme. □

Démonstration du lemme 3.31 . Notons $a = (x_1, x_2, \ldots, x_r)$ et $b = (y_1, y_2, \ldots, y_r)$. On calcule directement

$$\|a+b\|^2 = \sum_{i=0}^{r} (x_i+y_i)^2 = \|a\|^2 + \|b\|^2 + 2(a,b),$$

avec

$$(a,b) = \sum_{i=0}^{r} x_i y_i \leq \|a\|\cdot\|b\|,$$

d'après l'inégalité de Cauchy-Schwarz. Finalement

$$\|a+b\|^2 \leq \|a\|^2 + \|b\|^2 + 2\|a\|\cdot\|b\| = (\|a\|+\|b\|)^2.$$

Ceci montre la « première » inégalité triangulaire $\|a+b\| \leq \|a\| + \|b\|$. La deuxième se déduit de la première, exactement comme dans le corollaire 2.5. □

On en déduit enfin :

LEMME 3.33 – *Si $(u_n)_{n\geq 0}$ est une suite de vecteurs de $\mathbb{R}^r$ qui converge vers $\ell\in\mathbb{R}^r$, alors on a également*

$$\|u_n\| \xrightarrow[n\to\infty]{} \|\ell\|.$$

Démonstration. On utilise la deuxième inégalité triangulaire :

$$\big|\|\ell\| - \|u_n\|\big| \leq \|\ell - u_n\|.$$

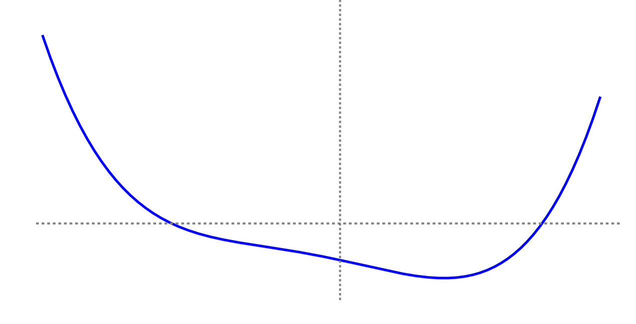
Le membre de droite tend vers 0 d'après le (2) de la proposition 3.29. □

voir l'exercice 2621

Chapitre 4

Continuité

Une fonction *continue*, intuitivement, est une fonction que l'on peut dessiner sans lever le stylo, comme celle-ci :



Ci-dessous, un dessin d'une fonction qui n'est pas continue. On a même l'intuition, plus précisément, qu'elle n'est pas continue au point x_0 :

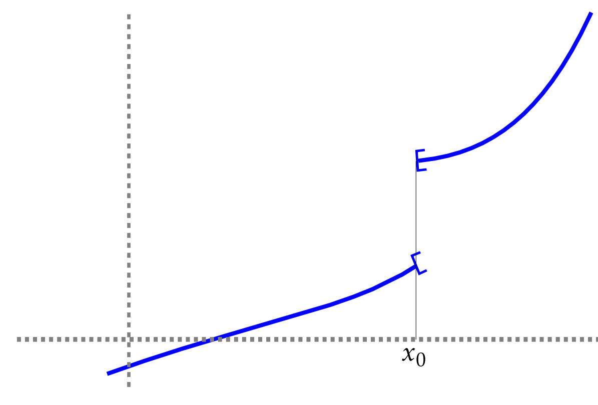


FIGURE 4.1 – *Le graphe d'une fonction qui n'est pas continue en x_0. Les crochets sont simplement là pour indiquer que la valeur de la fonction au point x_0 est celle indiquée sur la branche droite du graphe.*

Pourquoi s'intéresser aux fonctions continues ? La propriété cruciale d'une fonction continue est la suivante : puisque le graphe est tracé d'un seul tenant, alors si la fonction prend des valeurs positives et des valeurs négatives, on est sûr qu'elle prend également la valeur 0. Sur le premier graphe ci-dessus, la fonction est d'abord positive, puis prend quelques valeurs négatives ; et bien sûr elle passe par la valeur 0 (comment éviter cela si on ne peut pas lever le stylo ?). De même, puisque cette fonction reprend des valeurs positives, elle s'annule une deuxième fois.

En d'autres termes, si l'on sait qu'une fonction f est continue, alors on peut *prédire l'existence de solutions* de l'équation $f(x) = 0$. De nombreuses équations qui nous concernent peuvent se mettre sous cette forme, donc la notion de continuité va être très utile.

Oui mais comment traduire mathématiquement l'idée de continuité ? Il y a plusieurs façons de le faire, toutes assez abstraites au premier abord, et nous devrions avoir un seul critère pour juger du bien-fondé d'une définition : elle doit nous permettre de démontrer rigoureusement la propriété ci-dessus. Nous allons prendre la définition qui nous paraît la plus simple, et nous allons effectivement démontrer le célèbre « théorème des valeurs intermédiaires », qui en est la version précise.

DÉFINITION 4.2 – Soit $I \subset \mathbb{R}$, et soit $f : I \to \mathbb{R}$ une fonction. On dit que f est *continue* au point $x_0 \in I$ lorsque, pour toute suite $(u_n)_{n\geq 0}$ qui converge vers x_0, la suite $(f(u_n))_{n\geq 0}$ converge vers $f(x_0)$.

Lorsque f est continue en tout point $x_0 \in I$, on dit tout simplement qu'elle est continue sur I.

(La lettre I est pour « intervalle », puisque la plupart de nos exemples sont sur un intervalle, et certains théorèmes ne fonctionnent que dans ce cas ; ceci dit, en toute généralité I peut être n'importe quoi.)

Voyons un exemple :

LEMME 4.3 – *Soit $P \in \mathbb{R}[X]$ un polynôme. Alors la fonction $x \mapsto P(x)$ est continue sur $\mathbb{R}$.*

Démonstration. Soit donc une suite $(u_n)_{n\geq 0}$ telle que $u_n \to x_0$. Puisque P est un polynôme, la suite $(P(u_n))$ est obtenue à partir de (u_n) en faisant une série d'additions et de multiplications. Les limites de suites sont « compatibles » avec les sommes et les produits, comme la proposition 3.10 nous l'affirme, donc $P(u_n) \xrightarrow[n\to\infty]{} P(x_0)$. Ce qui signifie par définition que P est continue. □

EXEMPLE 4.4 – Le premier graphe de ce chapitre est celui de $x \mapsto x^4 + x^3 - x^2 - 5x + 1$. Dire que cette fonction est continue en x_0 revient à dire que, si $u_n \to x_0$, alors $u_n^4 + u_n^3 - u_n^2 - 5u_n + 1 \to x_0^4 + x_0^3 - x_0^2 - 5x_0 + 1$, ce qui est clair. Le dessin a été obtenu à l'aide d'un ordinateur, qui procède à toutes sortes d'approximations pendant le tracé, donc on ne peut pas conclure grand-chose de l'aspect de ce graphe. Toutefois, il est rassurant que le résultat ne soit pas contraire à notre intuition des fonctions continues.

EXEMPLE 4.5 – Voici un exemple de fonction qui n'est pas continue. Définissons f sur l'intervalle $[0,2]$ par :

$$f(x) = \begin{cases} 0 \text{ si } x < 1 \\ 1 \text{ si } x \geq 1 . \end{cases}$$

Le graphe de f fait donc un « saut » autour de la valeur 1. Pour montrer que f n'est pas continue, on va considérer la suite $u_n = 1 - \frac{1}{n}$. On a bien $u_n \to 1$, mais $f(u_n) = 0$ pour tout n, donc $(f(u_n))$ ne risque pas de converger vers $f(1) = 1$. Par définition, f n'est pas continue.

Avant d'énoncer ce théorème, un petit rappel sur les intervalles. Jusqu'à présent nous avons utilisé le mot « intervalle » en nous basant sur les souvenirs du lycée. Voici une définition simple :

Définition 4.6 – Soit $I \subset \mathbb{R}$. On dit que I est un *intervalle* lorsque, pour tous nombres a, b et c tels que $a < b < c$ avec $a \in I$ et $c \in I$, on a aussi $b \in I$.

Exemple 4.7 – On a les exemples suivants :

- Les intervalles *ouverts*, de la forme

$$]a,b[= \{x \mid a < x < b\},$$

avec $a, b \in \mathbb{R}$ ou même $a = -\infty$, $b = +\infty$.

- Les intervalles *fermés*, qui peuvent être « compacts », c'est-à-dire de la forme

$$[a,b] = \{x \mid a \leq x \leq b\} \quad \text{avec } a, b \in \mathbb{R},$$

ou bien « non-compacts », c'est-à-dire de la forme

$$[a,+\infty[\quad \text{ou} \quad]-\infty,b].$$

- Les intervalles *semi-ouverts*, qui sont de la forme

$$]a,b] \quad \text{ou} \quad [a,b[.$$

En fait, cette liste est complète. Vous montrerez à titre d'exercice que tout intervalle I est de l'un des types ci-dessus, et que de plus $a = \inf I$ et $b = \sup I$, lorsque l'une ou l'autre de ces bornes existe.

Voyons un contre-exemple. L'ensemble

$$I = [0,1] \cup [3,4]$$

n'est pas un intervalle. D'abord, il n'est pas dans la liste ci-dessus, et surtout on voit tout de suite que $1 \in I$, que $3 \in I$ mais que $2 \notin I$, ce qui contredit bien la définition.

Nous pouvons maintenant énoncer :

Théorème 4.8 (Théorème des valeurs intermédiaires) – *Soit I un intervalle, et soit $f : I \to \mathbb{R}$ une fonction continue. Soient $a < b$ deux éléments de I. Alors si y est un nombre quelconque compris entre $f(a)$ et $f(b)$, il existe (au moins) un nombre x avec $a \leq x \leq b$ tel que $f(x) = y$.*

(En d'autres termes, l'image d'un intervalle par une fonction continue est encore un intervalle.)

Démonstration. On va donner deux démonstrations, en commençant par un argument « avec des sup ». Supposons par exemple que $f(a) < f(b)$, de sorte que $f(a) \leq y \leq f(b)$. Posons

$$A = \{t \mid a \leq t \leq b \text{ et } f(t) \leq y\}.$$

C'est un ensemble non-vide ($a \in A$) et majoré (par b), donc il possède une borne supérieure $x = \sup A$.

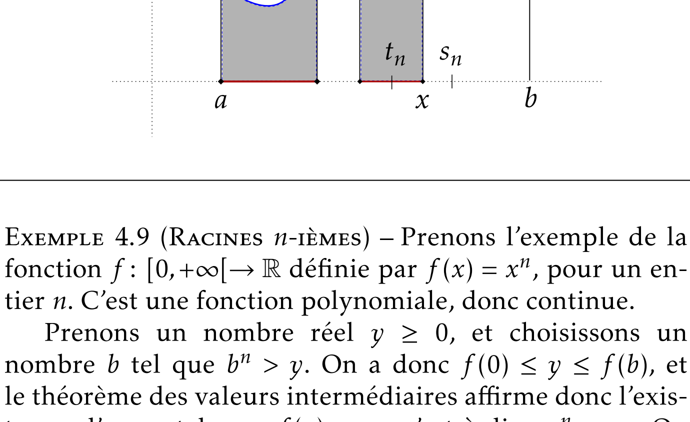
Soit n un entier ≥ 1. Considérons le nombre $x - \frac{1}{n} < x$. Ce n'est pas un majorant de A (puisque x est le plus petit), donc il existe $t_n \in A$ avec $x - \frac{1}{n} < t_n \leq x$. On a $t_n \to x$. La fonction f étant continue, on a également $f(t_n) \to f(x)$. Puisque $f(t_n) \leq y$, on a $f(x) \leq y$.

On va maintenant établir que $f(x) \geq y$, ce qui est d'ailleurs évident si $x = b$, et on se place donc dans le cas $x < b$. Le nombre $s_n = x + \frac{1}{n} > x$ ne peut pas appartenir à A, et $s_n \in [a,b]$ pour n suffisamment grand ; donc $f(s_n) > y$. On a $s_n \to x$ et, par continuité de f, il vient $f(s_n) \to f(x)$, d'où $f(x) \geq y$.

Voici maintenant une deuxième démonstration, à l'aide de suites monotones. On suppose toujours qu'on est dans le cas $f(a) < f(b)$, et on va construire par récurrence deux suites $(a_n)_{n\geq 0}$ et $(b_n)_{n\geq 0}$ telles que (a_n) est croissante, (b_n) est décroissante, $f(a_n) \leq y \leq f(b_n)$, et enfin $b_n - a_n \to 0$. À cette fin, on pose $a_0 = a$ et $b_0 = b$ pour commencer. Ensuite, en supposant que l'on a défini les nombres a_n et b_n pour un certain n, on pose $m = \frac{a_n + b_n}{2}$, et deux cas se présentent : ou bien $f(m) \leq y$ et on pose $a_{n+1} = m$, $b_{n+1} = b_n$, ou bien $f(m) > y$ et on pose $a_{n+1} = a_n$, $b_{n+1} = m$. Dans les deux cas, on a $b_{n+1} - a_{n+1} = \frac{b_n - a_n}{2}$ et en particulier $b_n - a_n = \frac{b_0 - a_0}{2^n} \to 0$. Toutes les propriétés annoncées sont alors évidemment vérifiées.

D'après le théorème 3.14, la suite (a_n) possède une limite ℓ_1, et donc $f(a_n) \to f(\ell_1)$ car f est supposée continue. L'inégalité $f(a_n) \leq y$ donne $f(\ell_1) \leq y$. De même la suite (b_n) possède une limite ℓ_2 et $f(\ell_2) \geq y$. Or $b_n - a_n \to 0 = \ell_2 - \ell_1$, d'où $\ell_1 = \ell_2$, et $f(\ell_1) = f(\ell_2) = y$. □

Notez bien que x n'est pas unique en général. Sur le dessin suivant on a représenté les éléments de la première démonstration, et on voit qu'on avait trois choix dans ce cas pour x, celui retenu étant le plus grand. La zone située entre la courbe et l'ensemble A a été grisée.



Exemple 4.9 (Racines n-ièmes) – Prenons l'exemple de la fonction $f : [0,+\infty[\to \mathbb{R}$ définie par $f(x) = x^n$, pour un entier n. C'est une fonction polynomiale, donc continue.

Prenons un nombre réel $y \geq 0$, et choisissons un nombre b tel que $b^n > y$. On a donc $f(0) \leq y \leq f(b)$, et le théorème des valeurs intermédiaires affirme donc l'existence d'un x tel que $f(x) = y$, c'est-à-dire $x^n = y$. On constate que tout nombre réel positif possède une racine n-ième.

De plus, si $0 \leq x_1 < x_2$, on a $x_1^n < x_2^n$; cette remarque simple entraîne l'unicité du $x \geq 0$ tel que $x^n = y$. La racine n-ième positive de y est bien définie, on la note $\sqrt[n]{y}$.

Nous avions démontré ce résultat pour $n = 2$, avec pas mal d'efforts (proposition 2.10). Le théorème des valeurs intermédiaires, maintenant qu'il est démontré, simplifie considérablement ce genre de questions.

Il nous reste à considérer les racines n-ièmes dans $\mathbb{C}$, ce qui nécessite de nouveaux outils.

Autres exemples de fonctions continues

Les fonctions « usuelles » sont toutes continues :

Proposition 4.10 – *Les fonctions suivantes sont continues en tout point de leur domaine de définition :*
- $x \mapsto e^x$ *sur* $\mathbb{R}$,
- $x \mapsto \sin(x)$ *sur* $\mathbb{R}$,
- $x \mapsto \cos(x)$ *sur* $\mathbb{R}$,
- $x \mapsto \tan(x)$ *sur* $\mathbb{R} \setminus \{\frac{\pi}{2} + k\pi \text{ avec } k \in \mathbb{Z}\}$,
- $x \mapsto \ln(x)$ *sur* $]0, +\infty[$,
- $x \mapsto \arcsin(x)$ *sur* $[-1, 1]$,
- $x \mapsto \arccos(x)$ *sur* $[-1, 1]$,
- $x \mapsto \arctan(x)$ *sur* $\mathbb{R}$,
- $x \mapsto \sqrt[n]{x}$ *sur* $[0, +\infty[$ *pour* $n \in \mathbb{N}$.

Pour l'instant, on ne risque pas de donner une démonstration de cette proposition : on n'a même pas de définition rigoureuse de la plupart de ces fonctions ! Toutefois, dans le reste de ce chapitre, on va établir que $x \mapsto \sqrt[n]{x}$ est continue, et nous verrons également qu'il suffit de montrer la continuité des quatre premières fonction dans la liste ci-dessus pour obtenir automatiquement la continuité des autres. Grâce à certaines formules de trigonométrie qu'il faudra établir et que vous devinez peut-être, on se ramènera à montrer seulement la continuité de l'exponentielle. Pour cela, on travaillera directement avec la définition donnée dans l'exemple 3.22. Nous traiterons ceci dans l'appendice intitulé « L'exponentielle ».

Pour construire encore plus de fonctions continues, on utilise le résultat suivant :

Proposition 4.11 – *Soient f et g deux fonctions définies sur* I *et continues en $x_0 \in$* I. *Alors*
- *(somme)* $x \mapsto f(x) + g(x)$ *est continue en* x_0,
- *(produit)* $x \mapsto f(x)g(x)$ *est continue en* x_0,
- *(inverse) si* $f(x) \neq 0$ *sur* I, *alors* $x \mapsto \frac{1}{f(x)}$ *est continue en* x_0.

En effet, ceci découle directement de la proposition 3.10.

Exemple 4.12 – Si l'on admet que le sinus et le cosinus sont des fonctions continues, alors

$$x \mapsto \tan(x) = \frac{\sin(x)}{\cos(x)}$$

est également continue là où elle est définie, c'est-à-dire là où le cosinus ne s'annule pas.

Par rapport aux suites, une nouveauté très simple :

Proposition 4.13 – *Soit $f : \mathrm{I} \to \mathrm{J}$ continue en $x_0 \in \mathrm{I}$, et soit $g : \mathrm{J} \to \mathbb{R}$ continue en $f(x_0) \in \mathrm{J}$. Alors la fonction $g \circ f : \mathrm{I} \to \mathbb{R}$, qui à x associe $g(f(x))$, est continue en x_0.*

Démonstration. Soit donc $(u_n)_{n \geq 0}$ une suite convergeant vers x_0. Posons $v_n = f(u_n)$. La suite (v_n) converge vers $f(x_0)$ car f est continue en x_0. La suite $(g(v_n))$ converge vers $g(f(x_0))$ car g est continue en $f(x_0)$. Donc $g \circ f$, par définition, est continue en x_0. □

Exemple 4.14 – Considérons une expression comme

$$x \mapsto \frac{e^{\cos(x)} - \ln(x)}{1 + (\arctan(x))^2}.$$

Les propositions ci-dessus permettent d'affirmer en un clin d'œil que cette fonction est continue sur $]0, +\infty[$. En effet, commençons par

$$x \mapsto 1 + (\arctan(x))^2 ;$$

la fonction arctan est continue (4.10), donc son carré aussi (4.11) ; la fonction constante égale à 1 est continue, donc la somme $1 + (\arctan(x))^2$ est continue (4.11 encore).

Ce dénominateur ne s'annule pas, donc

$$x \mapsto \frac{1}{1 + (\arctan(x))^2}$$

est continue (4.11).

Ensuite $x \mapsto e^{\cos(x)}$ est continue puisque c'est une composition de fonctions continues (4.13) ; l'expression $x \mapsto -\ln(x)$ est le produit du logarithme et de la fonction constante égale à -1, c'est donc une fonction continue. La somme des deux aussi : le numérateur est continu.

Enfin, toute l'expression de départ étant le produit de deux fonctions continues, il est continu.

Il faut s'entraîner à reconnaître très vite que ce genre d'expression donne une fonction continue.

voir les exercices 642, 649, 671, 645

Définition 4.15 – Soient $f : \mathrm{I} \to \mathbb{R}$ une fonction et $x_0 \in \mathbb{R}$, ou même $x_0 = \pm\infty$. On dit que f admet ℓ pour *limite* en x_0, et on note

$$\lim_{x \to x_0} f(x) = \ell,$$

lorsque pour toute suite $(u_n)_{n \geq 0}$ qui converge vers x_0, avec $u_n \in \mathrm{I}$, la suite $(f(u_n))$ converge vers ℓ.

Dans un premier temps, cette notion apparaît comme une reformulation de la continuité, notamment à cause du résultat suivant :

Proposition 4.16 – *Soient $f : \mathrm{I} \to \mathbb{R}$ et $x_0 \in \mathrm{I}$. Alors*

$$f \text{ est continue en } x_0 \iff f \text{ admet une limite en } x_0.$$

De plus, la limite est automatiquement $f(x_0)$.

Démonstration. Si f est continue en x_0, alors par définition

$$\lim_{x \to x_0} f(x) = f(x_0),$$

donc on a l'implication $\Rightarrow$. Pour montrer $\Leftarrow$, supposons que f admette la limite ℓ en x_0. Il suffit de prendre la suite constante $u_n = x_0$ pour constater que $(f(u_n))$ converge vers $f(x_0)$ (cette suite est elle-même constante). Donc $\ell = f(x_0)$. Il est alors clair que f est continue en x_0. □

Mais il ne faut pas s'y méprendre. Les limites apportent une souplesse nouvelle, puisque l'on *ne suppose pas que f est définie en x_0 dans la définition des limites*. Voyons des exemples.

Exemple 4.17 – Considérons la fonction $f :]0, +\infty[\to \mathbb{R}$ définie par $f(x) = \frac{1}{x}$.

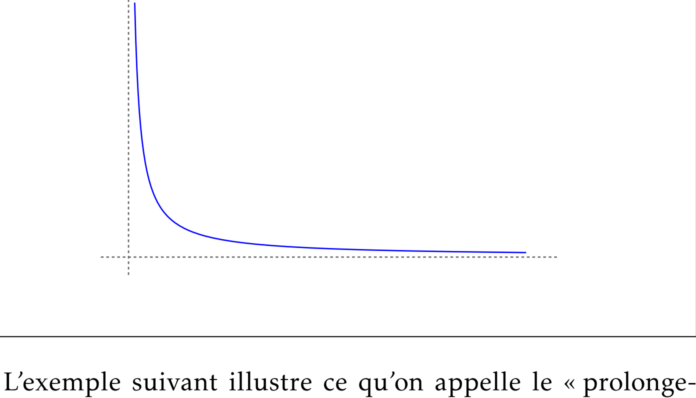
Regardons la limite en 0. Si $(u_n)_{n \geq 0}$ converge vers 0 avec u_n dans le domaine de définition de f, c'est-à-dire $u_n > 0$, on a $f(u_n) = \frac{1}{u_n} \to +\infty$. C'est donc que

$$\lim_{x \to 0} f(x) = +\infty.$$

Le même raisonnement donne

$$\lim_{x \to +\infty} f(x) = 0.$$

En fait le graphe a l'allure suivante :



L'exemple suivant illustre ce qu'on appelle le « prolongement par continuité ».

Exemple 4.18 – Soit $f : \mathbb{R} \smallsetminus \{0\} \to \mathbb{R}$ définie par

$$f(x) = x \sin\left(\frac{1}{x}\right).$$

Cette fonction est continue en tout $x_0 \neq 0$ comme on le voit facilement à partir des résultats ci-dessus.

Regardons la limite en 0, donc prenons (u_n) qui tend vers 0 avec $u_n \neq 0$. Alors $|f(u_n)| \leq |u_n|$ puisque $|\sin(\frac{1}{u_n})| \leq 1$. On en déduit $f(u_n) \to 0$, et donc

$$\lim_{x \to 0} f(x) = 0.$$

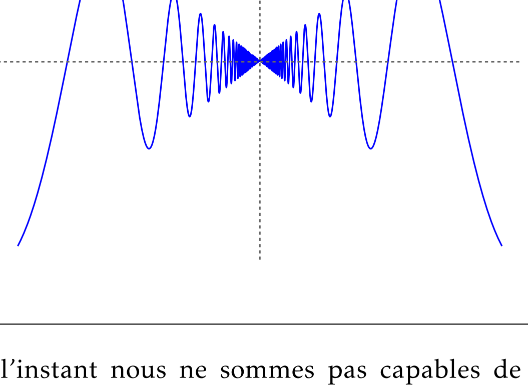
Définissons alors

$$\tilde{f}(x) = \begin{cases} f(x) \text{ si } x \neq 0, \\ 0 \text{ sinon.} \end{cases}$$

Les fonctions f et $\tilde{f}$ ont la même limite en tout point $x_0 \in \mathbb{R}$ (prenez le temps de vous en convaincre). De plus cette limite vaut toujours $\tilde{f}(x_0)$. On conclut que $\tilde{f}$ est continue.

On dit que $\tilde{f}$ prolonge f par continuité, puisqu'elle coïncide avec f sur le domaine de définition de cette dernière, et qu'elle est continue. Vue la limite de f en 0, on n'aurait pas pu prendre une autre valeur pour $\tilde{f}(0)$, donc le prolongement est unique.

Voici le graphe de f sur $[-0,2; 0,2]$:



Pour l'instant nous ne sommes pas capables de calculer beaucoup de limites (pas plus qu'au lycée). Dans le chapitre « Formules de Taylor » nous verrons une méthode rapide et facile, qui fonctionne dans beaucoup de cas.

voir les exercices 677, 675, 639, 646

Le résultat suivant est très utile. Il affirme que si une fonction continue satisfait une inégalité en un point, alors ceci reste vrai au « voisinage » du point :

PROPOSITION 4.19 – *Soit $f : \mathrm{I} \to \mathbb{R}$ une fonction continue en $x_0 \in \mathrm{I}$. On suppose que $f(x_0) > 0$. Alors il existe un intervalle ouvert $\mathrm{J} =]a, b[$ avec $x_0 \in \mathrm{J}$ tel que $f(x) > 0$ pour tout $x \in \mathrm{J} \cap \mathrm{I}$.*

Démonstration. On va même montrer qu'on peut prendre $a = x_0 - \frac{1}{n}$ et $b = x_0 + \frac{1}{n}$ pour un certain entier n. En effet, si ce n'était pas le cas, par l'absurde on trouverait pour chaque n un $x_n \in \mathrm{I}$ tel que

$$x_0 - \frac{1}{n} < x_n < x_0 + \frac{1}{n}$$

et tel que $f(x_n) \leq 0$. La suite (x_n) converge vers x_0, et par continuité de f la suite $(f(x_n))$ converge vers $f(x_0)$: on en déduit $f(x_0) \leq 0$ ce qui est absurde. □

En guise d'application nous allons montrer le théorème suivant, qui donne des définitions alternatives de la notion de continuité. Le (3) en particulier est utilisé dans de nombreux livres.

THÉORÈME 4.20 – *Soient $f : \mathrm{I} \to \mathbb{R}$ et $x_0 \in \mathrm{I}$. Les conditions suivantes sont équivalentes.*

1. *f est continue en x_0 ;*
2. *pour tout intervalle ouvert J contenant $f(x_0)$, il existe un intervalle ouvert I' contenant x_0 tel que $f(\mathrm{I}' \cap \mathrm{I}) \subset \mathrm{J}$;*
3. *pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que $|f(x) - f(x_0)| < \varepsilon$ pour tous les x tels que $|x - x_0| < \delta$.*

Démonstration. Montrons que (1) $\Rightarrow$ (2). L'intervalle J étant donné, il existe m et M tels que $f(x_0) \in]m, \mathrm{M}[\subset \mathrm{J}$. Soit alors $g(x) = f(x) - m$; c'est une fonction continue telle que $g(x_0) > 0$, donc d'après la proposition précédente on a aussi $g(x) > 0$ pour tous les $x \in \mathrm{I}$ dans un intervalle $]a_1, b_1[$. De même en considérant $h(x) = \mathrm{M} - f(x)$, on obtient $h(x) > 0$ pour $x \in \mathrm{I}$ dans un intervalle $]a_2, b_2[$. Sur l'intervalle $\mathrm{I}' =]a_1, b_1[\cap]a_2, b_2[$, on a $m < f(x) < \mathrm{M}$, donc $f(\mathrm{I}' \cap \mathrm{I}) \subset]m, \mathrm{M}[\subset \mathrm{J}$.

Montrons (2) $\Rightarrow$ (3). Prenons $\mathrm{J} =]f(x_0) - \varepsilon, f(x_0) + \varepsilon[$, alors le (2) donne un intervalle I', qui lui-même contient un intervalle de la forme $]x_0 - \delta, x_0 + \delta[$. L'inclusion $f(\mathrm{I}' \cap \mathrm{I}) \subset \mathrm{J}$ donne la conclusion du (3).

Montrons (3) $\Rightarrow$ (1). Soit donc $(u_n)_{n \geq 0}$ une suite d'éléments de I qui converge vers x_0 ; on doit montrer que $(f(u_n))$ converge vers $f(x_0)$. On prend donc $\varepsilon > 0$, et un δ comme dans le (3). Pour n suffisamment grand, on a $|u_n - x_0| < \delta$, et donc $|f(u_n) - f(x_0)| < \varepsilon$. □

voir l'exercice 670

– Fin de la première lecture –

Définition 4.21 – Une fonction f est dite *croissante* lorsque, pour tout x et y dans son domaine de définition, l'inégalité $x < y$ entraîne $f(x) \leq f(y)$. Elle est dite *décroissante* si $x < y$ entraîne au contraire $f(x) \geq f(y)$. (Ainsi f ne peut être à la fois croissante et décroissante que si elle est constante.)

On dit d'une fonction f qu'elle est *monotone* si elle est ou bien croissante, ou bien décroissante.

Enfin, on parle de fonction *strictement croissante* lorsque $x < y$ entraîne $f(x) < f(y)$ (inégalité stricte cette fois), et de même on a les concepts de fonction strictement décroissante et strictement monotone.

Nous verrons beaucoup d'exemples dans le chapitre sur la dérivabilité. Notons simplement que l'exponentielle est croissante, la fonction cosinus est décroissante sur $[0, \pi]$, et la même fonction cosinus vue comme une fonction définie sur $\mathbb{R}$ tout entier n'est pas monotone.

Le comportement des fonctions monotones vis-à-vis de la continuité est particulièrement simple. Commençons par une sorte de réciproque au théorème des valeurs intermédiaires :

Proposition 4.22 – *Soient* I *un intervalle et* $f\colon I \to \mathbb{R}$ *une fonction monotone. Alors* f *est continue si et seulement si* $f(I)$ *est un intervalle.*

Par exemple, la fonction sur la figure 4.1 est croissante. Elle n'est pas continue, et son ensemble image est en deux morceaux.

Démonstration. Le théorème des valeurs intermédiaires affirme que si f est continue, alors $f(I)$ est un intervalle. Supposons que f est croissante et montrons la réciproque (le cas où f est décroissante est similaire). Prenons $x_0 \in I$ et supposons pour l'instant que $f(x_0)$ n'est pas une borne de l'intervalle $f(I)$.

Soit J un intervalle ouvert contenant $f(x_0)$. Alors $J \cap f(I)$ est un intervalle contenant $f(x_0)$, donc contenant un intervalle $[m, M]$ avec $m < f(x_0) < M$. Par définition $m = f(a)$ et $M = f(b)$ pour $a, b \in I$. Comme f est croissante, on a $a < x_0 < b$ d'une part, et d'autre part pour $a < x < b$ on a $m \leq f(x) \leq M$. En posant $I' =]a, b[\subset I$, on a en particulier $f(I') \subset J$. D'après le (2) du théorème 4.20, ceci montre que f est continue en x_0.

Lorsque $f(x_0)$ est une borne de $f(I)$, on a $m = f(x_0)$ ou $M = f(x_0)$, et selon le cas, $f(x_0)$ est un minimum ou un maximum de la fonction croissante f. On peut adapter facilement la démonstration (laissé en exercice). □

Dans le reste de ce chapitre, nous aurons besoin des notions de fonction injective, surjective, et bijective (définitions 1.6, 1.10, 1.13).

Proposition 4.23 – *Soient* I *un intervalle, et* $f\colon I \to \mathbb{R}$ *une fonction continue. Alors* f *est strictement monotone si et seulement si elle est injective.*

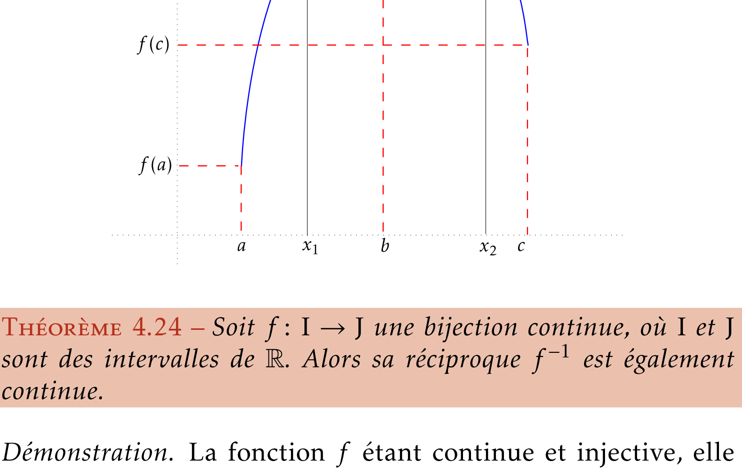
Démonstration. Si f est strictement croissante (disons), alors en prenant deux éléments $x_1 \neq x_2 \in I$ on doit avoir $x_1 < x_2$ ou $x_2 < x_1$, donc $f(x_1) < f(x_2)$ ou $f(x_2) < f(x_1)$ selon le cas, et certainement $f(x_1) \neq f(x_2)$. Donc f est injective.

Voyons la réciproque. Supposons que f est continue et injective, et prenons $a \in I$. On va montrer que f est strictement monotone sur $]a, +\infty[\cap I$; comme a est arbitraire, on aura bien établi que f est strictement monotone.

Soit $g(x) = f(x) - f(a)$, que l'on voit comme une fonction définie sur l'intervalle $I \cap]a, +\infty[$. Par injectivité de f, la fonction g ne s'annule pas. D'après le théorème des valeurs intermédiaires, elle ne peut pas changer de signe, donc disons par exemple que l'on a $g(x) > 0$ pour tous les $x > a$, c'est-à-dire $f(x) > f(a)$. Dans ce cas on va montrer que f est strictement croissante sur $]a, +\infty[\cap I$.

En effet, si ce n'était pas le cas, on aurait deux valeurs b et c avec $a < b < c$ telles que $f(b) > f(c)$ (inégalité stricte par injectivité de f). Mais alors, prenons n'importe quelle valeur y telle que $f(c) < y < f(b)$ et $f(a) < y < f(b)$. En appliquant le théorème des valeurs intermédiaires sur l'intervalle $[a, b]$, on trouve $x_1 < b$ tel que $f(x_1) = y$. En faisant de même sur $[b, c]$ on trouve $x_2 > b$ tel que $f(x_2) = y$. Ceci est absurde puisque f est injective. □

La fin de la démonstration est illustrée sur la figure suivante.



Théorème 4.24 – *Soit* $f\colon I \to J$ *une bijection continue, où* I *et* J *sont des intervalles de* $\mathbb{R}$. *Alors sa réciproque* f^{-1} *est également continue.*

Démonstration. La fonction f étant continue et injective, elle est strictement monotone par la proposition précédente. Donc f^{-1} aussi. De plus $f^{-1}(J) = I$, qui est un intervalle par hypothèse, donc f^{-1} est continue d'après la proposition 4.22. □

Exemple 4.25 – Lorsque nous aurons (enfin) montré que la fonction exponentielle est continue, nous déduirons du théorème ci-dessus que sa réciproque le logarithme est également continue. De même les fonctions arccosinus, arcsinus, et arctangente sont continues parce que les fonctions cosinus, sinus, et tangentes sont continues, comme nous le montrerons.

Pour l'instant, nous pouvons déjà établir fermement que $x \mapsto \sqrt[n]{x}$ est continue sur $[0, +\infty[$: en effet, c'est la réciproque de la fonction $x \mapsto x^n$, qui est continue puisqu'elle est polynomiale.

Puisque la continuité s'exprime en termes de convergence de suites, et que nous savons ce que signie « converger » pour une suite de vecteurs (voir proposition 3.29), nous pouvons étendre sans problème la définition principale de ce chapitre :

Définition 4.26 – Soient $X \subset \mathbb{R}^n$, et $f : X \to \mathbb{R}^m$ une fonction. On dit que f est *continue* en $x \in X$ lorsque pour toute suite $(u_n)_{n\geq 0}$ d'éléments de X qui converge vers x (dans $\mathbb{R}^n$), la suite $(f(u_n))$ converge vers $f(x)$ (dans $\mathbb{R}^m$).

On dit que f admet ℓ pour limite en $x_0 \in \mathbb{R}^n$ lorsque pour toute suite $(u_n)_{n\geq 0}$ qui converge vers x_0, avec chaque $u_n \in X$, la suite $(f(u_n))$ converge vers ℓ (ceci même si $x_0 \notin X$).

Notons qu'une telle fonction est de la forme

$$(x_1, x_2, \ldots, x_n) \mapsto (f_1(x_1, \ldots x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

Chaque fonction f_i est définie sur X et prend ses valeurs dans $\mathbb{R}$. On appelle ces fonctions les « composantes » de f.

Le résultat suivant se démontre exactement comme les énoncés correspondants pour les fonctions d'une seule variable.

Proposition 4.27 – *Les sommes, produits et inverses de fonctions continues, lorsqu'elles sont définies, sont continues. La composition de deux fonctions continues est encore continue.*

Avec les notations ci-dessus, la fonction f est continue si et seulement si chaque composante f_i est continue.

De plus, f est continue en $x \in X$ si et seulement si elle admet la limite $f(x)$ en ce point.

Enfin, une fonction f est continue en x_0 si et seulement si, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que $\|f(x) - f(x_0)\| < \varepsilon$ pour tous les $x \in X$ tels que $\|x - x_0\| < \delta$.

Attention, par contre la réciproque d'une fonction continue de plusieurs variables n'est pas toujours continue, contrairement au cas traité dans le théorème 4.24.

Les ouverts

Soit $U \subset \mathbb{R}^n$. On dit que U est *ouvert dans* $\mathbb{R}^n$ lorsque, pour tout $x \in U$, on peut trouver un $r > 0$ tel que la « boule de centre x et de rayon r », c'est-à-dire l'ensemble

$$B(x, r) = \{y \in \mathbb{R}^n \mid \|x - y\| < r\},$$

vérifie $B(x, r) \subset U$. Par exemple,

$$U = \{(x, y) \mid x > 0 \text{ et } y > 0\}$$

est ouvert dans $\mathbb{R}^2$. D'ailleurs un intervalle de $\mathbb{R}$ est « ouvert dans $\mathbb{R}$ » exactement lorsque c'est un intervalle ouvert au sens élémentaire.

Toute partie de $\mathbb{R}^n$ de la forme $F = \mathbb{R}^n \smallsetminus U$, où U est ouvert dans $\mathbb{R}^n$, est dite *fermée dans* $\mathbb{R}^n$. Par exemple

$$F = \{(x, y) \mid x \geq 0 \text{ et } y \geq 0\}$$

est fermée dans $\mathbb{R}^2$. Dans un tout premier temps, on peut forger son intuition en se rappelant que les inégalités strictes définissent des ouverts et que les inégalités larges définissent des fermés.

Vous pourrez vous amuser à montrer que F est fermée dans $\mathbb{R}^n$ si et seulement si, pour toute suite (u_n) d'éléments de F telle que $u_n \to \ell \in \mathbb{R}^n$, on a en fait $\ell \in F$ (« une partie est fermée exactement lorsqu'elle contient ses limites »).

Avec ces définitions, on peut établir la chose suivante (essayez) : une fonction $f : \mathbb{R}^m \to \mathbb{R}^n$ est continue $\Longleftrightarrow$ pour tout U ouvert dans $\mathbb{R}^n$, la partie $f^{-1}(U)$ est ouverte dans $\mathbb{R}^m$. (Et pareil avec « fermée » au lieu de « ouverte » d'ailleurs.)

C'est la définition de la continuité la plus flexible, qui va se généraliser très bien : pour pouvoir parler de continuité, il suffit de pouvoir parler de parties ouvertes. Par exemple, pour $X \subset \mathbb{R}^n$, on dit que les parties ouvertes de X sont celles de la forme $U \cap X$ avec U ouverte dans $\mathbb{R}^n$; avec cette notion, on retombe bien sur la définition de la continuité donnée dans ce chapitre, comme vous pouvez le montrer (en y consacrant pas mal d'efforts tout de même).

En deuxième année, vous étudierez en détails les *espaces métriques*, qui sont en gros des ensembles sur lesquels on peut mesurer la distance entre deux éléments. De ce fait on peut définir les boules, comme ci-dessus, et donc les ouverts.

En troisième année vous étudierez les ouverts eux-mêmes. C'est une définition un peu sèche d'un sujet ludique qu'on appelle la *topologie*.

Exemple 4.28 – La *projection* p_i, définie sur $\mathbb{R}^n$ par

$$p_i(x_1, x_2, \ldots, x_n) = x_i,$$

est continue : on le vérifie directement à partir des définitions.

Partant de là, on peut utiliser des sommes et produits, par exemple

$$(x_1, x_2, x_3) \mapsto 2x_1x_3 - x_2^5$$

est continue sur $\mathbb{R}^3$ d'après la proposition précédente.

On peut aussi composer avec des fonctions usuelles :

$$(x_1, x_2) \mapsto \sin(x_1x_2)$$

est continue, ainsi que

$$(x_1, x_2, x_3) \mapsto e^{x_1^2 - x_3} - \arctan(x_2 - 1).$$

voir l'exercice 1794

Exemple 4.29 – Voici un exemple plus sophistiqué. On va identifier l'ensemble des matrices $M_n(\mathbb{R})$ avec $\mathbb{R}^{n^2}$ pour toutes les questions de continuité (le fait de disposer les nombres en tableau ne change rien à l'affaire). De même on va identifier $M_n(\mathbb{R}) \times M_n(\mathbb{R})$ avec $\mathbb{R}^{2n^2}$.

Ceci étant fait, il est légitime de demander si la fonction suivante est continue :

$$\begin{array}{rccc} f : & M_n(\mathbb{R}) \times M_n(\mathbb{R}) & \longrightarrow & M_n(\mathbb{R}) \\ & (A, B) & \mapsto & f(A, B) = AB. \end{array}$$

Et la réponse est oui : si $A = (a_{ij})_{i,j}$ et $B = (b_{ij})_{i,j}$, alors sur la ligne i, dans la colonne j de $f(A, B)$ on trouve

$$\sum_{k=0}^{n} a_{i,k}\, b_{k,j}.$$

Cette expression est continue (elle est obtenue à partir des projections en faisant des produits et des sommes). Puisque les composantes de f sont continues, c'est que f est elle-même continue.

Notons maintenant $GL_n(\mathbb{R})$ l'ensemble des matrices *inversibles* de $M_n(\mathbb{R})$; c'est une notation standard qui fait référence à l'expression « groupe linéaire ». Que dire de la continuité de la fonction suivante ?

$$\begin{array}{rccc} g : & GL_n(\mathbb{R}) & \longrightarrow & GL_n(\mathbb{R}) \\ & A & \mapsto & A^{-1}. \end{array}$$

C'est loin d'être une question abstraite ou inutile. Lorsque vous confiez à un ordinateur la tâche de calculer l'inverse d'une matrice A à coefficients réels, dans de nombreux cas vous allez entrer une approximation B de la matrice A (disons en ne donnant qu'une dizaine de chiffres après la virgule). L'ordinateur vous donne la valeur de B^{-1}. Est-ce que, du fait que B était proche de A, on peut s'attendre à ce que B^{-1} soit proche de A^{-1} ? C'est ce qu'on demande lorsque la continuité de g est étudiée.

Dans le cas $n = 2$, nous pouvons répondre : en effet d'après la proposition 12.9, la fonction g s'écrit dans ce cas

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(La même proposition affirme que sur $GL_2(\mathbb{R})$, la quantité $ad - bc$ ne s'annule pas.) Cette expression est visiblement continue.

Nous allons voir que g est continue pour tout n, mais pour cela il va nous falloir développer la théorie des *déterminants*, dans le chapitre du même nom.

Chapitre 5

Bolzano-Weierstrass

Définition 5.1 – Soit $(u_n)_{n\geq 0}$ une suite. Une *sous-suite* de (u_n) est une suite de la forme $(u_{\sigma(n)})_{n\geq 0}$ où $\sigma \colon \mathbb{N} \to \mathbb{N}$ est une fonction strictement croissante.

Les exemples typiques sont $(u_{2n})_{n\geq 0}$ et $(u_{2n+1})_{n\geq 0}$.

Théorème 5.2 (Bolzano & Weierstrass) – *Soit $(u_n)_{n\geq 0}$ une suite de nombre réels. On suppose qu'il existe deux nombres a et b tels que $u_n \in [a,b]$ pour chaque indice n. Alors il existe une sous-suite $(u_{\sigma(n)})_{n\geq 0}$ qui possède une limite $\ell \in [a,b]$.*

Souvent on énonce : « de toute suite de réels bornée on peut extraire une sous-suite convergente ».

Démonstration. Posons $a_0 = a$, $b_0 = b$, et $m = \frac{a_0+b_0}{2}$, le milieu de $[a_0, b_0]$. Soit $A \subset \mathbb{N}$ l'ensemble des entiers n tels que $u_n \in [a_0, m]$, et soit B l'ensemble des entiers n tels que $u_n \in [m, b_0]$.

Les ensembles A et B ne peuvent pas être tous les deux finis, puisque $A \cup B = \mathbb{N}$. Si A est infini, on pose $a_1 = a_0$ et $b_1 = m$; dans le cas contraire on pose $a_1 = m$ et $b_1 = b_0$.

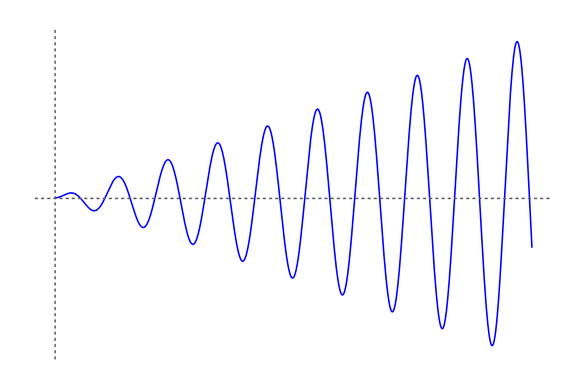
Par récurrence, on construit a_{n+1} et b_{n+1} à partir de a_n et b_n de la même manière, en s'arrangeant pour qu'il y ait une infinité de termes de la suite dans l'intervalle $[a_{n+1}, b_{n+1}]$.

On finit avec deux suites (a_n) et (b_n) telles que $|b_n - a_n| = \frac{|b-a|}{2^n}$ (puisque l'on divise la longueur de l'intervalle par 2 à chaque étape), et en particulier $b_n - a_n \to 0$. De plus (a_n) est croissante et majorée, donc converge vers ℓ_1, alors que (b_n) est décroissante et minorée et donc converge vers ℓ_2. On en conclut que $\ell_1 = \ell_2$.

Pour chaque entier n, choisissons maintenant un entier $\sigma(n)$ tel que $a_n \leq u_{\sigma(n)} \leq b_n$ en s'arrangeant pour que σ soit croissante (c'est possible par construction). Il est clair que $u_{\sigma(n)} \to \ell_1 = \ell_2$. □

Les applications sont plus théoriques que pratiques, au moins dans un premier temps. Elles sont par contre fondamentales, et vont prendre de plus en plus d'importance au fur et à mesure de vos études en mathématiques.

Nous avons vu avec le théorème des valeurs intermédiaires (4.8) que l'image d'un intervalle par une fonction continue est encore un intervalle. Il y a plusieurs types d'intervalles : ouverts, fermés, semi-ouverts... Est-ce que l'image d'un intervalle est du même type que celui-ci ? La réponse est non, comme sur la figure ci-dessous.



Ici on voit une portion du graphe d'une fonction continue définie sur l'intervalle fermé $[0,+\infty[$, et son image est visiblement l'intervalle ouvert $]-\infty,+\infty[$ (il reste à imaginer la suite du graphe, évidemment). Méfiance donc.

Par contre on a le résultat suivant, qui est notre première application de Bolzano & Weierstrass.

Proposition 5.3 – *Soit f une fonction continue définie sur l'intervalle compact $I = [a,b]$. Alors $f(I)$ est aussi un intervalle compact.*

Démonstration. Soit $J = f(I)$, on sait que c'est un intervalle d'après le théorème des valeurs intermédiaires. Soit $m = \inf(J)$ ou $m = -\infty$ si l'inf n'existe pas ; de même soit $M = \sup(J)$ ou $M = +\infty$ si le sup n'existe pas. On a donc $J = (m, M)$ où les parenthèses signifient qu'on ne sait pas encore s'il s'agit de [ou].

Prenons une suite (y_n) telle que $y_n \to M$, avec $y_n \in J$. Par définition on a $y_n = f(x_n)$ pour un certain $x_n \in I = [a,b]$. D'après le théorème de Bolzano et Weierstrass, on peut « extraire » une sous-suite $(x_{\sigma(n)})$ qui converge vers $\ell \in [a,b]$. Par continuité de f, on a $f(x_{\sigma(n)}) = y_{\sigma(n)} \to f(\ell)$. Comme $(y_{\sigma(n)})$ est une sous-suite de (y_n), elle doit converger vers M, et donc $M = f(\ell)$.

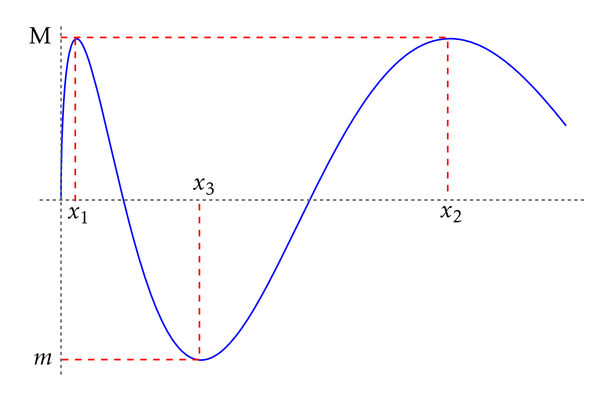
On en conclut que $M \neq +\infty$, et que $M \in J$. On procède de même pour montrer que $m \neq -\infty$ et que $m \in J$. Finalement $J = [m, M]$. □

Notons bien, en particulier, que l'on a le résultat suivant :

Corollaire 5.4 – *Soit f une fonction continue définie sur un intervalle compact. Alors f atteint son maximum et son minimum.*

L'expression « f atteint son maximum et son minimum » contient plusieurs choses, qui sont toutes des conséquences du fait que l'image de f est de la forme $[m, M]$, mais qu'il est bon d'énoncer séparément. Tout d'abord le sup des valeurs prises par la fonction, que l'on note M, n'est pas $+\infty$; mais plus précisément, on sait aussi qu'il existe un x dans l'ensemble de définition tel que $f(x) = M$. C'est pour cette raison que l'on parle du maximum de f et pas seulement de son sup, pour insister. De même avec le minimum m.

Il peut y avoir plusieurs valeurs pour lesquelles les extrema sont atteints, bien sûr. Sur la figure ci-dessous, le maximum de la fonction est atteint en x_1 et x_2, et le minimum est atteint en x_3.



On comparera la situation avec celle de la fonction $x \mapsto \frac{1}{x}$, définie sur l'intervalle $]0,+\infty[$ (qui n'est pas compact) : elle ne possède ni maximum ni minimum. L'inf des valeurs de cette fonction est 0, mais cette valeur n'est pas atteinte.

Le corollaire prédit donc l'existence de valeurs maximales et minimales sous des hypothèses assez simples. Il est bien utile, comme on va le voir.

– Fin de la première lecture –

Les arguments utilisant le théorème de Bolzano et Weierstrass sont tellement efficaces que l'on en vient à donner un nom aux ensembles sur lesquels on peut l'adapter.

DÉFINITION 5.5 – Soit $X \subset \mathbb{R}^n$. On dit que X est *compact* lorsque de toute suite $(u_n)_{n\geq 0}$ avec $u_n \in X$ on peut extraire une sous-suite $(u_{\sigma(n)})_{n\geq 0}$ qui converge vers $\ell \in X$.

Avec ce vocabulaire, le théorème de Bolzano et Weierstrass peut être interprété comme affirmant que les ensembles que nous avions d'ores et déjà appelés « intervalles compacts » sont effectivement compacts au sens de cette définition. Ce sont d'ailleurs les seuls intervalles ayant cette propriété. Par exemple $[0,+\infty[$ n'est pas compact, puisque la suite $(2^n)_{n\geq 0}$ n'a pas de sous-suite convergente ; de même $]0,1[$ n'est pas compact, puisque la suite $(\frac{1}{n})_{n\geq 1}$ ne possède que des sous-suites qui convergent vers 0, et $0 \notin]0,1[$ (noter la condition $\ell \in X$, très importante, dans la définition).

voir l'exercice 1763

L'étude des compacts se fera plus en détails en deuxième voire troisième année. Nous allons cependant voir un exemple riche de conséquences.

PROPOSITION 5.6 – *Soit* $R \subset \mathbb{R}^2$ *un rectangle de la forme*

$$R = [a,b] \times [c,d].$$

Alors R *est compact.*

Démonstration. Soit $(u_n)_{n\geq 0}$ une suite d'éléments de R, et notons $u_n = (x_n, y_n)$.

D'après le théorème de Bolzano et Weierstrass, on peut trouver une sous-suite $(x_{\sigma(n)})$ qui converge vers $\ell \in [a,b]$. Appliquons le même théorème à la suite (z_n) définie par $z_n = y_{\sigma(n)}$: il existe une sous-suite $(z_{\tau(n)})$ qui converge vers $\ell' \in [c,d]$. Notons que $z_{\tau(n)} = y_{\sigma(\tau(n))}$.

Considérons la suite $(x_{\sigma(\tau(n))})$: c'est une sous-suite de $(x_{\sigma(n)})$ donc elle converge vers ℓ. Donc finalement $u_{\sigma(\tau(n))}$ converge vers (ℓ, ℓ'). □

Citons un exemple de résultat dont la démonstration se déduit immédiatement de celle que nous avons donnée pour les intervalles compacts.

PROPOSITION 5.7 – *Soit* $f : X \to \mathbb{R}$ *une fonction continue, où* $X \subset \mathbb{R}^n$ *est un compact. Alors* f *atteint son maximum et son minimum.*

À titre d'exercice vous montrerez ceci en adaptant l'argument donné pour le corollaire 5.4 : vous verrez qu'il n'y a essentiellement rien à changer.

Pour montrer l'existence d'un maximum ou d'un minimum d'une fonction f dont on ne sait pas grand'chose, ou avec laquelle on ne souhaite pas faire de calculs compliqués (comme les dérivées du chapitre suivant), on essaie souvent de se ramener au corollaire 5.4. Lorsque la fonction en question n'est pas définie sur un compact, il faut faire des efforts supplémentaires. Voici un exemple simple.

PROPOSITION 5.8 – *Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction continue.*

1. *Supposons que pour toute suite $(u_n)_{n\geq 0}$ telle que $\|u_n\| \to +\infty$, on a $f(u_n) \to +\infty$. Alors f atteint son minimum sur $\mathbb{R}^2$.*
2. *Supposons que pour toute suite $(u_n)_{n\geq 0}$ telle que $\|u_n\| \to +\infty$, on a $f(u_n) \to 0$. Alors f atteint son maximum sur $\mathbb{R}^2$.*

Démonstration. Montrons le (1), le (2) étant similaire. Soit

$$R_n = [-n, n] \times [-n, n].$$

Ce rectangle est compact d'après la proposition 5.6, donc sur R_n la fonction f atteint un minimum en u_n (proposition 5.7).

Puisque $R_n \subset R_{n+1}$, on a $f(u_{n+1}) \leq f(u_n)$ (le minimum diminue quand on le prend sur une partie plus grande). Montrons maintenant qu'il existe un entier N tel que tous les termes u_n appartiennent à R_N. Par l'absurde, si ce n'était pas le cas, on aurait une sous-suite $(u_{\sigma(n)})_{n\geq 0}$ telle que $\|u_{\sigma(n)}\| \to +\infty$: il suffit de prendre $u_{\sigma(n)}$ à l'extérieur du rectange R_n. Par hypothèse $f(u_{\sigma(n)}) \to +\infty$, ce qui est absurde puisque $(f(u_{\sigma(n)}))$ est décroissante. Donc N existe.

Si maintenant $x \in \mathbb{R}^2$, on a $x \in R_n$ pour un certain n, et donc $f(x) \geq f(u_n)$; mais $u_n \in R_N$, donc $f(u_n) \geq f(u_N)$. Finalement $f(u_N)$ est le minimum de f. □

La prochaine application du théorème de Bolzano et Weierstrass concerne une propriété fine des fonctions continues. Rappelons que dans le (3) du théorème 4.20, nous avons montré qu'une fonction f était continue au point x_0 si et seulement si la condition suivante était satisfaite : pour tout $\varepsilon > 0$ on doit pouvoir trouver un $\delta > 0$ de telle sorte que $|f(x) - f(x_0)| < \varepsilon$ dès que $|x - x_0| < \delta$. Voyons ça sur un exemple.

EXEMPLE 5.9 – Prenons $f(x) = \frac{1}{x}$ sur l'intervalle $]0,1]$, et soit x_0 dans cet intervalle. Montrons que f est continue en x_0 directement à partir de la définition (c'est-à-dire sans utiliser le fait qu'il s'agit de l'inverse d'une fonction continue qui ne s'annule pas). Soit donc $\varepsilon > 0$. Calculons :

$$|f(x) - f(x_0)| = \left|\frac{1}{x} - \frac{1}{x_0}\right| = \left|\frac{x - x_0}{x x_0}\right|.$$

On a donc $|f(x) - f(x_0)| < \varepsilon$ dès que $|x - x_0| < \varepsilon x x_0$. Mais on ne peut certainement pas prendre $\delta = \varepsilon x x_0$, puisque δ ne doit pas dépendre de x, par définition.

Par contre, choisissons un nombre $\rho > 0$ tel que $\rho < x_0$, et prenons $\delta' = \varepsilon\rho^2$. Pour tous les x dans l'intervalle $[\rho, 1]$ on a $\delta' < \varepsilon x x_0$, donc l'inégalité $|x - x_0| < \delta'$ entraîne bien $|f(x) - f(x_0)| < \varepsilon$.

Pour finir, nous devons prendre δ tel que $0 < \delta \leq \delta'$ d'une part, et $\delta < x_0 - \rho$ d'autre part. Avec ce δ, l'inégalité $|x - x_0| < \delta$ entraîne $x = x_0 - (x_0 - x) > x_0 - \delta > \rho$, c'est-à-dire que $x \in [\rho, 1]$; l'argument ci-dessus donne donc $|f(x) - f(x_0)| < \varepsilon$ dès que $|x - x_0| < \delta$.

Une chose à retenir de ce calcul, c'est que sur un intervalle de la forme $[\rho, 1]$ on peut prendre *le même* δ' pour tous les points x_0 à la fois (à savoir $\delta' = \varepsilon\rho^2$). Alors que sur $]0,1]$, nous venons de le voir, il faut choisir un δ qui dépend de x_0.

Ce phénomène porte un nom :

DÉFINITION 5.10 – Soit $I \subset \mathbb{R}$, et soit $f : I \to \mathbb{R}$. On dit que f est *uniformément continue* sur I lorsque pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que $|f(x) - f(x_0)| < \varepsilon$ dès que l'on choisit $x, x_0 \in I$ tels que $|x - x_0| < \delta$.

Avec ce langage, nous pouvons dire que la fonction $x \mapsto \frac{1}{x}$ est uniformément continue sur chaque intervalle compact de la forme $[\rho, 1]$ pour $0 < \rho < 1$. Par contre, sur l'intervalle $]0,1]$, la fonction définie par la même formule n'est pas uniformément continue. C'est l'essence de l'exemple 5.9.

On peut se demander pourquoi rentrer dans des considérations si précises. La réponse arrivera avec le chapitre sur les intégrales : il se trouve que, pour définir rigoureusement l'intégrale d'une fonction, nous aurons besoin de savoir que ladite fonction est uniformément continue. Heureusement, nous n'aurons pas à refaire un travail comme ci-dessus pour chaque fonction, puisque le théorème suivant nous épargne toutes les difficultés.

THÉORÈME 5.11 (HEINE) – *Soit f une fonction continue sur un intervalle compact. Alors elle est uniformément continue.*

Démonstration. Procédons par l'absurde. Si f n'est pas uniformément continue, alors il existe un $\varepsilon > 0$ tel que pour chaque $\delta > 0$, on peut choisir x et y tels que $|x - y| < \delta$ et cependant $|f(x) - f(y)| \geq \varepsilon$. Prenons $\delta_n = \frac{1}{n}$ pour chaque entier $n \geq 1$, et notons x_n et y_n nos choix pour δ_n.

Soit $I = [a, b]$ l'intervalle sur lequel f est définie. D'après la proposition 5.6, le rectangle $R = I \times I$ est compact. La suite définie par $u_n = (x_n, y_n)$ possède donc une sous-suite $u_{\sigma(n)} = (x_{\sigma(n)}, y_{\sigma(n)})$ qui converge vers $(\ell, \ell') \in R$. En d'autres termes on a $x_{\sigma(n)} \to \ell$ et $y_{\sigma(n)} \to \ell'$, et donc $x_{\sigma(n)} - y_{\sigma(n)} \to \ell - \ell'$. De plus, puisque

$$\left|x_{\sigma(n)} - y_{\sigma(n)}\right| < \delta_{\sigma(n)} = \frac{1}{\sigma(n)} \xrightarrow[n\to\infty]{} 0,$$

on constate que $\ell = \ell'$.

Enfin, par continuité de f, on a $f(x_{\sigma(n)}) \to f(\ell)$ et $f(y_{\sigma(n)}) \to f(\ell)$, donc

$$f(x_{\sigma(n)}) - f(y_{\sigma(n)}) \to 0.$$

Mais dans la mesure où

$$\left|f(x_{\sigma(n)}) - f(y_{\sigma(n)})\right| \geq \varepsilon > 0,$$

cette dernière convergence vers 0 est impossible. Cette conclusion absurde montre que f est uniformément continue. □

Chapitre 6

Dérivées

DÉFINITION 6.1 – Soit f une fonction définie sur un intervalle I. On dit que f est *dérivable* au point $x_0 \in \mathrm{I}$ lorsque le *taux d'accroissement*, c'est-à-dire la fonction définie par

$$\mathrm{T}_{x_0}(x) = \frac{f(x) - f(x_0)}{x - x_0}$$

(pour $x \neq x_0$) possède une limite (finie). Lorsque c'est le cas, cette limite est notée $f'(x_0)$, et ce nombre est appelé le *nombre dérivé* en x_0.

La fonction $x \mapsto f'(x)$, lorsqu'elle est définie en tout point de I, est appelée la *dérivée* de f.

Le taux d'accroissement $\mathrm{T}_{x_0}(x)$ ci-dessus est la « pente » de la droite qui passe par $(x_0, f(x_0))$ et par $(x, f(x))$. En faisant varier x, avec x_0 fixé, on obtient toute une famille de droites. Lorsque f est dérivable en x_0, ces droites atteignent une position « limite » lorsque x se rapproche de x_0. La droite passant par $(x_0, f(x_0))$ et dont la pente est cette valeur limite $f'(x_0)$ est appelée la *tangente* au graphe de f au point x_0. La situation est illustrée sur le dessin suivant.

Lorsque $f'(x_0) > 0$, la droite tangente est le graphe d'une fonction croissante ; il est raisonnable de penser alors à f elle-même comme à une fonction croissante au voisinage de x_0, et dans ce chapitre nous allons rendre cette intuition précise. Puisque vous connaissez le résultat (sans démonstration !) depuis le lycée, inutile de vous faire patienter : nous verrons que lorsque $f'(x_0) > 0$ pour tous les x_0, alors la fonction est croissante.

Le plus étonnant avec les dérivées reste la facilité avec laquelle on peut les calculer.

EXEMPLE 6.2 – Prenons $f(x) = x$, sur $\mathbb{R}$ tout entier par exemple. Pour tout x_0, on a

$$\mathrm{T}_{x_0}(x) = \frac{x - x_0}{x - x_0} = 1\,.$$

Cette fonction a donc certainement une limite, qui vaut 1. C'est-à-dire que $f'(x_0) = 1$, pour tout $x_0 \in \mathbb{R}$.

Autre exemple simple, si $g(x) = c$ est une fonction constante, alors le taux d'accroissement est

$$\frac{c - c}{x - x_0} = 0\,,$$

et donc $g'(x_0) = 0$.

Un peu plus compliqué, prenons $h(x) = \frac{1}{x}$ sur $\mathbb{R}^*$. Alors le taux d'accroissement est

$$\frac{h(x) - h(x_0)}{x - x_0} = \frac{1}{x - x_0}\left(\frac{1}{x} - \frac{1}{x_0}\right) = -\frac{1}{xx_0}\,.$$

Cette expression tend vers $-\frac{1}{x_0^2}$ lorsque $x \to x_0$, donc h est dérivable et $h'(x_0) = -\frac{1}{x_0^2}$.

Avant même de donner d'autres exemples, notons la chose suivante :

LEMME 6.3 – *Soit f une fonction dérivable en x_0. Alors f est également continue en x_0.*

Démonstration. On écrit simplement

$$f(x) - f(x_0) = (x - x_0) \cdot \frac{f(x) - f(x_0)}{x - x_0} \longrightarrow 0 \times f'(x_0) = 0\,,$$

donc f admet pour limite $f(x_0)$ lorsque $x \to x_0$, ce qui signifie bien qu'elle est continue. □

EXEMPLE 6.4 – Donnons quelques exemples de fonctions qui ne sont pas dérivables. Le plus simple est de prendre une fonction qui n'est pas continue : d'après le lemme, elle ne peut pas être dérivable non plus.

Mais il existe des fonctions continues qui ne sont pas dérivables. Prenons par exemple la « valeur absolue », c'est-à-dire la fonction $x \mapsto |x|$, définie sur $\mathbb{R}$. Prenons $x_0 = 0$ et examinons le taux d'accroissement :

$$\mathrm{T}_0(x) = \frac{|x| - |0|}{x - 0} = \begin{cases} 1 \text{ si } x > 0\,, \\ -1 \text{ si } x < 0\,. \end{cases}$$

Cette expression n'a pas de limite en 0, puisque par exemple on a $\mathrm{T}_0(\frac{1}{n}) \xrightarrow[n\to\infty]{} 1$ alors que $\mathrm{T}_0(-\frac{1}{n}) \xrightarrow[n\to\infty]{} -1$. Donc la fonction valeur absolue n'est pas dérivable en 0. (Et en x_0, pour $x_0 \neq 0$?)

Autre exemple, la fonction définie sur $[0, +\infty[$ par $x \mapsto \sqrt{x}$. En $x_0 = 0$, le taux d'accroissement est

$$\mathrm{T}_0(x) = \frac{\sqrt{x} - \sqrt{0}}{x - 0} = \frac{1}{\sqrt{x}}\,,$$

défini sur $]0, +\infty[$. On a donc $\mathrm{T}_0(x) \to +\infty$ lorsque $x \to 0$, et il n'y a pas de limite finie : la fonction n'est pas dérivable en 0 (et ailleurs ?).

PROPOSITION 6.5 – *Soient f et g deux fonctions définies sur I et dérivables en $x_0 \in \mathrm{I}$. Alors*
- ◇ *(somme) $x \mapsto f(x) + g(x)$ est dérivable en x_0, et sa dérivée en ce point est $f'(x_0) + g'(x_0)$.*
- ◇ *(produit) $x \mapsto f(x)g(x)$ est dérivable en x_0, et sa dérivée en ce point est $f'(x_0)g(x_0) + f(x_0)g'(x_0)$.*

Démonstration. Montrons la formule pour le produit (celle pour la somme est facile et laissée en exercice). On étudie le taux d'accroissement :

$$\frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0} = f(x)\left(\frac{g(x) - g(x_0)}{x - x_0}\right) + g(x_0)\left(\frac{f(x) - f(x_0)}{x - x_0}\right).$$

Puisque f est continue en x_0 en vertu du lemme précédent, cette expression a bien pour limite $f(x_0)g'(x_0) + f'(x_0)g(x_0)$, comme annoncé. □

EXEMPLE 6.6 – Prenons $f(x) = g(x) = x$. Alors la proposition indique que la dérivée de $x \mapsto x^2$ est $x \mapsto 1 \times x + x \times 1 = 2x$. Continuons : la dérivée de $x \mapsto x^3 = x^2 \times x$ est, toujours d'après la formule sur le produit, donnée par $x \mapsto (2x) \times x + x^2 \times 1 = 3x^2$.

En continuant de cette manière, on montre par récurrence (faites-le) que la dérivée de $x \mapsto x^n$ est $x \mapsto nx^{n-1}$.

Si on se donne une constante c, et que l'on applique encore et toujours la formule pour le produit, on constate que la dérivée de $x \mapsto cx^n$ est $x \mapsto 0 \times x^n + c \times nx^{n-1} = cnx^{n-1}$.

Enfin, grâce à la formule pour la somme, on constate que toute fonction polynomiale, c'est-à-dire de la forme

$$x \mapsto a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

est dérivable, de dérivée

$$x \mapsto a_1 + 2a_2 x + 3a_3 x^2 + \cdots + na_n x^{n-1}\,.$$

Voici une autre définition de la dérivabilité, qui va paraître artificielle pour l'instant mais dont on comprendra le caractère naturel dans le chapitre sur les formules de Taylor.

LEMME 6.7 – *f est dérivable en x_0 si et seulement s'il existe deux nombres a_0 et a_1 tels qu'on peut écrire*

$$f(x_0 + h) = a_0 + a_1 h + h\varepsilon(h)\,,$$

où $\varepsilon(h) \to 0$ quand $h \to 0$. De plus, lorsque ces nombres existent, on a $a_0 = f(x_0)$ et $a_1 = f'(x_0)$.

Lisez soigneusement la démonstration ci-dessous, qui est particulièrement simple, afin de comprendre qu'il s'agit juste d'une reformulation des choses. Géométriquement, le lemme affirme que la fonction $h \mapsto f(x_0) + f'(x_0)h$, dont le graphe est une droite (la droite tangente), est une « bonne » approximation de la fonction $h \mapsto f(x_0 + h)$, puisque la différence entre les deux, qui vaut $h\varepsilon(h)$, est le produit de *deux* fonctions qui tendent vers 0.

Démonstration. Supposons d'abord que f est dérivable en x_0, et posons (nous n'avons guère le choix)

$$\varepsilon(h) = \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0)\,,$$

pour $h \neq 0$, et $\varepsilon(0) = 0$. Par définition, on a bien $\varepsilon(h) \to 0$ lorsque $h \to 0$, et on a tout fait pour que $f(x_0 + h) = f(x_0) + f'(x_0)h + h\varepsilon(h)$.

Montrons la réciproque ; supposons que a_0 et a_1 existent, tels que $f(x_0 + h) = a_0 + a_1 h + h\varepsilon(h)$ avec $\varepsilon(h) \to 0$. Tout d'abord en faisant tendre h vers 0, on constate que

$$\lim_{x \to x_0} f(x) = \lim_{h \to 0} f(x_0 + h) = a_0\,.$$

Mais alors, la proposition 4.16 nous dit que $a_0 = f(x_0)$.

Par suite le taux d'accroissement en x_0 vaut

$$\frac{f(x) - f(x_0)}{x - x_0} = \frac{f(x_0 + h) - f(x_0)}{h} = a_1 + \varepsilon(h)\,,$$

en posant $h = x - x_0$. Lorsque $x \to x_0$, ce taux d'accroissement tend donc vers a_1, ce qui par définition signifie que f est dérivable en x_0 et que $f'(x_0) = a_1$. □

Avec cette nouvelle formulation, il devient facile de montrer un résultat sur la composition des fonctions :

PROPOSITION 6.8 – *Soit $g\colon \mathrm{I} \to \mathrm{J}$ une fonction dérivable en $x_0 \in \mathrm{I}$, et soit $f\colon \mathrm{J} \to \mathbb{R}$ une fonction dérivable en $g(x_0) \in \mathrm{J}$. Alors $f \circ g$ est dérivable en x_0, et de plus*

$$(f \circ g)'(x_0) = f'(g(x_0))\,g'(x_0)\,.$$

Démonstration. D'après le lemme, on peut écrire

$$g(x_0 + h) = g(x_0) + g'(x_0)\,h + h\varepsilon(h)\,,$$

avec $\varepsilon(h) \to 0$ lorsque $h \to 0$, et de même en posant $y_0 = g(x_0)$,

$$f(y_0 + u) = f(y_0) + f'(y_0)\,u + u\phi(u)\,,$$

avec $\phi(u) \to 0$ lorsque $u \to 0$. On commence donc par écrire

$$\begin{aligned} f \circ g\,(x_0 + h) &= f\,[g(x_0 + h)] \\ &= f\,[g(x_0) + g'(x_0)\,h + h\varepsilon(h)] \\ &= f(y_0 + u)\,, \end{aligned}$$

en posant $u = g'(x_0)\,h + h\varepsilon(h)$. On peut donc poursuivre :

$$\begin{aligned} f \circ g\,(x_0 + h) &= f(y_0) + f'(y_0)\,u + u\phi(u) \\ &= f(y_0) + f'(y_0)\,[g'(x_0)h + h\varepsilon(h)] + u\phi(u) \\ &= f(y_0) + f'(y_0)g'(x_0)h + h\theta(h)\,, \end{aligned}$$

où l'on a rassemblé tous les termes manquants dans $\theta(h)$, c'est-à-dire que l'on a posé

$$\theta(h) = f'(y_0)\varepsilon(h) + (g'(x_0) + \varepsilon(h))\phi\,[h(g'(x_0) + \varepsilon(h))]\,.$$

Cette expression est peut-être compliquée, mais l'on retiendra simplement que $\theta(h) \to 0$ lorsque $h \to 0$.

On donc trouvé $a_0 = f(y_0)$ et $a_1 = f'(y_0)g'(x_0)$ comme dans le lemme, et on conclut que $f \circ g$ est dérivable en x_0 comme annoncé. □

EXEMPLE 6.9 – Prenons $f(t) = \frac{1}{t}$ sur $\mathrm{J} = \mathbb{R}^*$. Nous avons vu dans l'exemple 6.2 que $f'(t) = -\frac{1}{t^2}$.

Si on se donne maintenant une fonction g définie sur I telle que $g(x) \neq 0$ pour tout $x \in \mathrm{I}$, alors on peut la voir comme une fonction $\mathrm{I} \to \mathrm{J}$. On peut donc considérer $f \circ g$, et on a tout simplement $f \circ g\,(x) = \frac{1}{g(x)}$.

La proposition affirme que cette fonction est dérivable, et que sa dérivée en x est

$$f'(g(x))\,g'(x) = -\frac{1}{g(x)^2}\,g'(x) = -\frac{g'(x)}{g(x)^2}\,.$$

Ce résultat est à savoir, donc nous allons l'énoncer séparément.

PROPOSITION 6.10 – *Soit $g\colon \mathrm{I} \to \mathbb{R}$ une fonction dérivable qui ne s'annule pas. Alors la fonction $x \mapsto \frac{1}{g(x)}$ est dérivable, de dérivée $x \mapsto -\frac{g'(x)}{g(x)^2}$.*

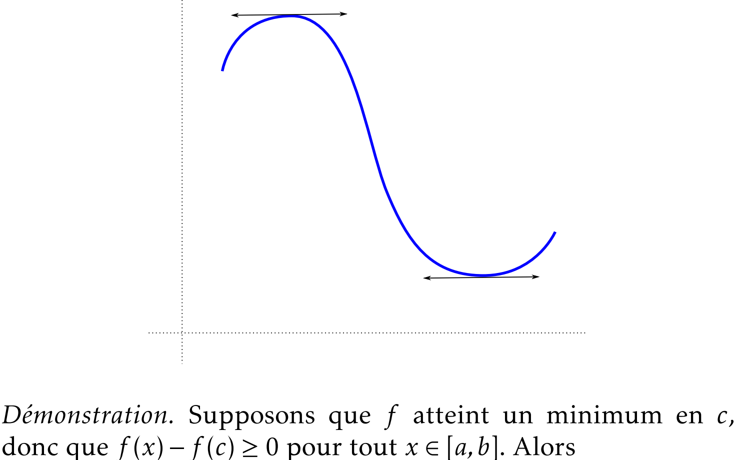
PROPOSITION 6.11 – *Les fonctions suivantes sont dérivables sur le domaine indiqué :*
- ◇ *$x \mapsto e^x$ sur $\mathbb{R}$, et sa dérivée est $x \mapsto e^x$,*
- ◇ *$x \mapsto \sin(x)$ sur $\mathbb{R}$, et sa dérivée est $x \mapsto \cos(x)$,*
- ◇ *$x \mapsto \cos(x)$ sur $\mathbb{R}$, et sa dérivée est $x \mapsto -\sin(x)$,*
- ◇ *$x \mapsto \tan(x)$ sur $\mathbb{R} \setminus \{\frac{\pi}{2} + k\pi \text{ avec } k \in \mathbb{Z}\}$, et sa dérivée est $x \mapsto 1 + (\tan(x))^2$,*
- ◇ *$x \mapsto \ln(x)$ sur $]0, +\infty[$, et sa dérivée est $x \mapsto \frac{1}{x}$,*
- ◇ *$x \mapsto \arcsin(x)$ sur $]-1, 1[$, et sa dérivée est $x \mapsto \frac{1}{\sqrt{1-x^2}}$,*
- ◇ *$x \mapsto \arccos(x)$ sur $]-1, 1[$, et sa dérivée est $x \mapsto -\frac{1}{\sqrt{1-x^2}}$,*
- ◇ *$x \mapsto \arctan(x)$ sur $\mathbb{R}$, et sa dérivée est $x \mapsto \frac{1}{1+x^2}$,*
- ◇ *$x \mapsto \sqrt[n]{x} = x^{1/n}$ sur $]0, +\infty[$ pour $n \geq 1$, et sa dérivée est $x \mapsto \frac{1}{n}(\sqrt[n]{x})^{1-n} = \frac{1}{n}x^{1/n - 1}$.*

Ce n'est pas la première fois que nous sommes dans cette situation : nous n'avons pas de (vraie) définition des fonctions exponentielle, sinus et cosinus, donc aucun espoir de faire cette démonstration pour l'instant. Dans l'appendice intitulé « L'exponentielle », nous pourrons remédier à cela.

On peut par contre calculer la dérivée de la fonction tangente à l'aide des formules pour cosinus et sinus : faites-le. Dans la deuxième partie de ce chapitre, nous verrons un résultat sur la dérivée de la réciproque d'une fonction, à partir duquel nous pourrons démontrer les formules ci-dessus pour les cinq derniers exemples.

Le théorème des accroissements finis est aux fonctions dérivables ce que le théorème des valeurs intermédiaires est aux fonctions continues : c'est ce résultat qui justifie les définitions.
Commençons par une remarque simple et importante :

LEMME 6.12 – *Soit $f\colon [a,b] \to \mathbb{R}$ une fonction. On suppose qu'il existe $c \in]a,b[$ tel que f est dérivable en c, et tel que f atteint un extremum (maximum ou minimum) en c. Alors $f'(c) = 0$.*



Démonstration. Supposons que f atteint un minimum en c, donc que $f(x) - f(c) \geq 0$ pour tout $x \in [a,b]$. Alors

$$\frac{f(x)-f(c)}{x-c} \geq 0 \text{ pour } x > c\,.$$

En passant à la limite, on obtient $f'(c) \geq 0$. Mais

$$\frac{f(x)-f(c)}{x-c} \leq 0 \text{ pour } x < c\,,$$

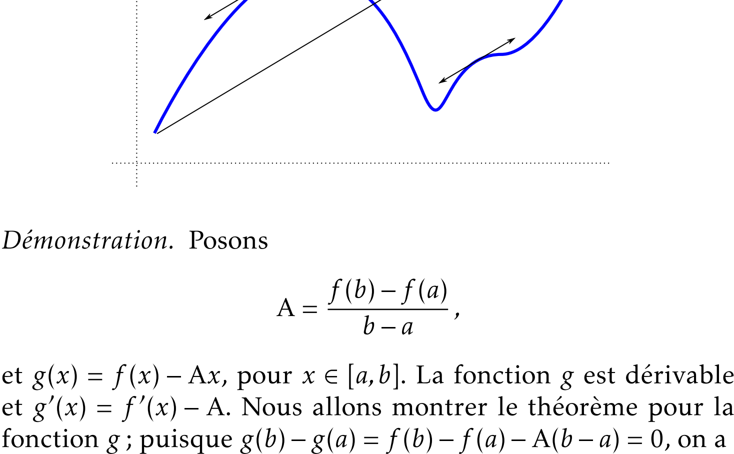
et en passant à la limite on obtient $f'(c) \leq 0$.
Finalement $f'(c) = 0$, comme annoncé. On procède de manière similaire dans le cas d'un maximum. □

Attention à ne pas faire dire à ce lemme plus que ce qu'il ne dit vraiment. Tout d'abord il ne faut pas oublier la condition $c \in]a,b[$: penser à la fonction $f(x) = x$ sur $[0,1]$, elle atteint un minimum en 0 et un maximum en 1, mais sa dérivée ne s'annule pas. Ensuite, noter que la réciproque du lemme est fausse : par exemple la fonction $f(x) = x^3$ sur $\mathbb{R}$ vérifie $f'(0) = 0$, mais on ne peut trouver aucun intervalle $[a,b]$ comme dans l'énoncé.

THÉORÈME 6.13 (ACCROISSEMENTS FINIS) – *Soit $f\colon [a,b] \to \mathbb{R}$ une fonction continue. On suppose que f est dérivable sur $]a,b[$.*
Alors il existe un nombre c avec $a < c < b$ tel que

$$f'(c) = \frac{f(b)-f(a)}{b-a}\,.$$

Bien sûr le nombre c n'est pas unique en général.



Démonstration. Posons

$$\mathrm{A} = \frac{f(b)-f(a)}{b-a}\,,$$

et $g(x) = f(x) - \mathrm{A}x$, pour $x \in [a,b]$. La fonction g est dérivable et $g'(x) = f'(x) - \mathrm{A}$. Nous allons montrer le théorème pour la fonction g ; puisque $g(b) - g(a) = f(b) - f(a) - \mathrm{A}(b-a) = 0$, on a

$$\frac{g(b)-g(a)}{b-a} = 0\,,$$

et il s'agit de montrer qu'il existe c tel que $g'(c) = 0$. Mais alors $g'(c) = f'(c) - \mathrm{A} = 0$ et on a bien $f'(c) = \mathrm{A}$, donc le théorème est montré pour f également.
(La fonction g vérifie $g(b) = g(a)$, et dans ce cas particulier le théorème s'appelle « théorème de Rolle ».)
D'après la proposition 5.3, on a $g([a,b]) = [m,\mathrm{M}]$, puisque g est continue. Si g est constante, alors $g'(x) = 0$ pour tout x, et il n'y a rien à montrer ; on peut donc supposer que $m < \mathrm{M}$. Soit $x_1 \in [a,b]$ tel que $g(x_1) = m$ et soit x_2 tel que $g(x_2) = \mathrm{M}$.
Si $x_1 \in]a,b[$, on prend $c = x_1$. Puisque g atteint alors un minimum en c, on a $g'(c) = 0$ d'après le lemme.
Si au contraire $x_1 = a$ ou b, on a $g(a) = g(b) = m$. Mais alors $x_2 \in]a,b[$ puisque $m \neq \mathrm{M}$. Dans ce cas on pose $c = x_2$, et puisque g atteint un maximum en c, le même lemme donne encore $g'(c) = 0$. □

La plus importante conséquence est bien sûr :

COROLLAIRE 6.14 – *Soit f une fonction dérivable définie sur l'intervalle* I. *Alors*
- ◇ *si $f'(x) \geq 0$ pour tous les $x \in \mathrm{I}$, alors la fonction f est croissante ;*
- ◇ *si $f'(x) > 0$ pour tous les $x \in \mathrm{I}$, alors f est strictement croissante ;*
- ◇ *si $f'(x) \leq 0$, resp. $f'(x) < 0$, alors f est décroissante, resp. strictement décroissante ;*
- ◇ *si $f'(x) = 0$ pour tous les $x \in \mathrm{I}$, alors f est constante.*

Démonstration. Montrons le premier point, les autres étant similaires (le dernier est même une conséquence des autres). Supposons donc que $f'(x) \geq 0$ pour $x \in \mathrm{I}$. Si $a < b$ sont deux éléments de I, alors d'après le théorème il existe c tel que

$$\frac{f(b)-f(a)}{b-a} = f'(c) \geq 0\,.$$

On en déduit que $f(b) \geq f(a)$. □

Vous connaissez le principe depuis la terminale, mais un petit rappel ne peut pas faire de mal : l'intérêt des dérivées réside dans le fait que les calculs sont assez faciles, et en tout cas se ramènent à un ensemble de règles « formelles » que l'on peut même apprendre à un ordinateur. Et pourtant, grâce à ce dernier corollaire, les informations que l'on obtient sur les fonctions sont utiles et précises.
Au lycée vous avez normalement passé beaucoup de temps à étudier des fonctions à l'aide de ce résultat, ce qui nécessite un certain entraînement. Nous allons à présent donner un exemple archi-simple pour rappeler un peu la méthode ; dans les exercices vous êtes invités à vous refaire la main. Dans le chapitre suivant nous verrons quelques exemples supplémentaires.

EXEMPLE 6.15 – Regardons la fonction f définie sur $\mathbb{R}$ par

$$f(x) = ax^2 + bx + c\,,$$

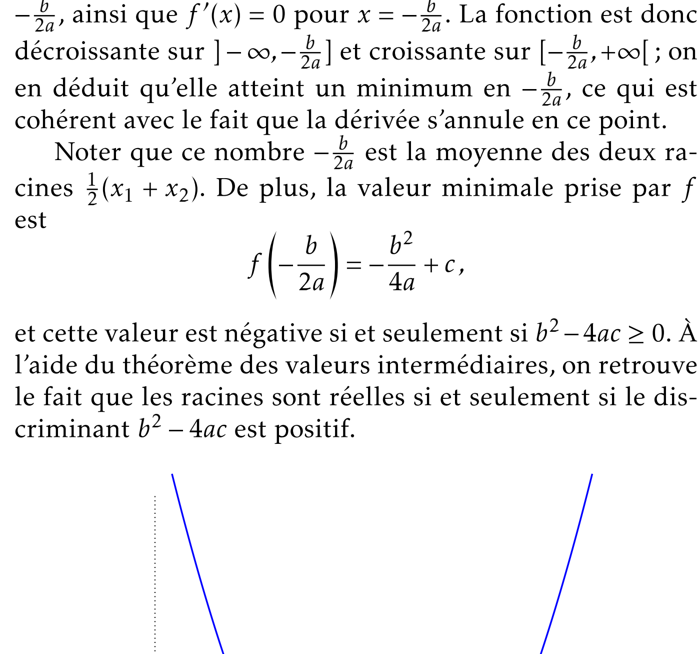
où $a, b, c \in \mathbb{R}$, avec $a \neq 0$. Soit $\delta \in \mathbb{C}$ un nombre tel que $\delta^2 = b^2 - 4ac$; nous savons que les racines de f, qui peuvent être réelles ou complexes, sont

$$x_1 = \frac{-b-\delta}{2a} \qquad \text{et} \qquad x_2 = \frac{-b+\delta}{2a}\,.$$

Regardons la dérivée : $f'(x) = 2ax + b$. Supposons que $a > 0$. On a donc $f'(x) > 0$ pour $x > -\frac{b}{2a}$, et $f'(x) < 0$ pour $x < -\frac{b}{2a}$, ainsi que $f'(x) = 0$ pour $x = -\frac{b}{2a}$. La fonction est donc décroissante sur $]-\infty, -\frac{b}{2a}]$ et croissante sur $[-\frac{b}{2a}, +\infty[$; on en déduit qu'elle atteint un minimum en $-\frac{b}{2a}$, ce qui est cohérent avec le fait que la dérivée s'annule en ce point.
Noter que ce nombre $-\frac{b}{2a}$ est la moyenne des deux racines $\frac{1}{2}(x_1 + x_2)$. De plus, la valeur minimale prise par f est

$$f\left(-\frac{b}{2a}\right) = -\frac{b^2}{4a} + c\,,$$

et cette valeur est négative si et seulement si $b^2 - 4ac \geq 0$. À l'aide du théorème des valeurs intermédiaires, on retrouve le fait que les racines sont réelles si et seulement si le discriminant $b^2 - 4ac$ est positif.



voir les exercices 698, 699, 700, 701, 703, 705, 709, 1221, 1223, 715, 717, 718, 721

– Fin de la première lecture –

Définition 6.16 – Soit $f : \mathrm{I} \to \mathbb{R}$ une fonction et soit $k \geq 0$. On dit que f est *k-lipschitzienne* sur I lorsque

$$|f(x) - f(y)| \leq k|x - y|,$$

pour tous $x, y \in \mathrm{I}$.

Lorsque $0 < k < 1$, on dit parfois d'une fonction lipschitzienne qu'elle est « contractante », c'est-à-dire qu'elle réduit les distances.

Vous montrerez à titre d'exercice très facile que si f est k-lipschitzienne, alors elle est continue, et même uniformément continue (définition 5.10).

Il est important de garder en tête que lorsque f est dérivable, alors cette notion nouvelle se ramène à un critère très simple :

Lemme 6.17 – *Soit $f : \mathrm{I} \to \mathbb{R}$ une fonction dérivable. Alors f est k-lipschitzienne si et seulement si $|f'(x)| \leq k$ pour tout $x \in \mathrm{I}$.*

Pour l'intuition il est donc raisonnable, dès lors qu'on a affaire à une fonction lipschitzienne, de penser à une fonction dérivable de dérivée bornée.

Démonstration. Si f est k-lipschitzienne, on écrit

$$\left|\frac{f(x) - f(x_0)}{x - x_0}\right| \leq k\left|\frac{x - x_0}{x - x_0}\right| = k,$$

d'où $|f'(x_0)| \leq k$ en passant à la limite.

Réciproquement si $|f'(x_0)| \leq k$ pour chaque x_0, alors on utilise le théorème des accroissements finis pour écrire

$$\left|\frac{f(x) - f(y)}{x - y}\right| = |f'(c)| \leq k$$

pour tous x, y. Le résultat en découle. □

Théorème 6.18 (Théorème du point fixe) – *Soit $f : \mathrm{I} \to \mathrm{I}$ une fonction k-lipschitzienne, pour $0 < k < 1$. Alors f possède un unique « point fixe » dans I, c'est-à-dire qu'il existe un unique $x_0 \in \mathrm{I}$ tel que $f(x_0) = x_0$.*

De plus, si $u_0 \in \mathrm{I}$ est choisi arbitrairement, et si l'on définit une suite $(u_n)_{n \geq 0}$ par $u_{n+1} = f(u_n)$, alors $(u_n) \xrightarrow[n \to \infty]{} x_0$.

Démonstration. Commençons par l'unicité. Si x_0 et x_1 sont deux points fixes de f, alors $|f(x_1) - f(x_0)| = |x_1 - x_0|$ d'une part, mais par hypothèse on a aussi $|f(x_1) - f(x_0)| \leq k|x_1 - x_0|$. On a donc $|x_1 - x_0| = 0$ puisque $k < 1$, d'où $x_1 = x_0$.

Prenons maintenant u_0 et définissons (u_n) comme dans l'énoncé. Si on peut montrer que (u_n) possède une limite ℓ, alors la suite $(f(u_n))$ doit converger vers $f(\ell)$ par continuité de f ; mais $f(u_n) = u_{n+1}$, donc $(f(u_n))$ est une sous-suite de (u_n), et à ce titre elle converge vers ℓ. Donc $\ell = f(\ell)$ et par unicité, $\ell = x_0$.

Il faut donc montrer que (u_n) converge. Écrivons

$$u_n = u_0 + (u_1 - u_0) + (u_2 - u_1) + \cdots + (u_n - u_{n-1}).$$

En d'autres termes en posant $a_n = u_n - u_{n-1}$ on a

$$u_n = u_0 + \sum_{i=1}^{n} a_i \,.$$

Nous allons montrer que la série de terme général a_i converge absolument, et donc converge d'après le théorème 3.21. Pour cela, notons que

$$\begin{aligned}
|a_i| = |u_i - u_{i-1}| &= |f(u_{i-1}) - f(u_{i-2})| \\
&\leq k|u_{i-1} - u_{i-2}| = k|f(u_{i-2}) - f(u_{i-3})| \\
&\leq k^2|u_{i-2} - u_{i-3}| \\
&\leq k^3|u_{i-3} - u_{i-4}| \\
&\leq \cdots \\
&\leq k^{i-1}|u_1 - u_0|.
\end{aligned}$$

Ceci montre que

$$\sum_{i=1}^{n}|a_i| \leq |u_1 - u_0|(1 + k + k^2 + \cdots + k^{n-1}) = |u_1 - u_0|\frac{1 - k^n}{1 - k} \leq |u_1 - u_0|\frac{1}{1 - k}.$$

On a donc bien

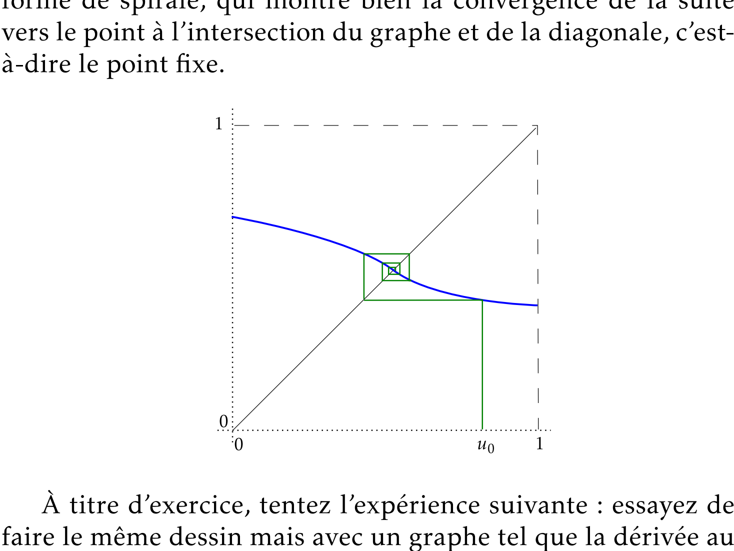
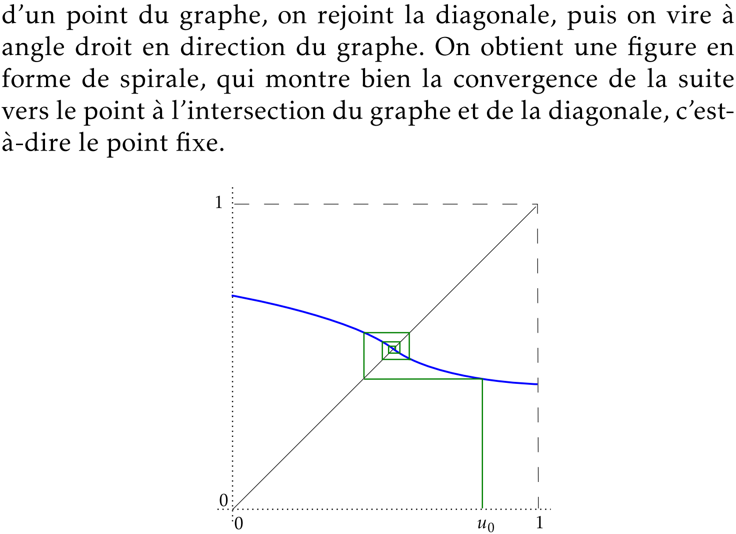
$$\sum_{i=1}^{+\infty}|a_i| < +\infty,$$

c'est-à-dire que la série est absolument convergente. □

Il est très facile de représenter la situation sur un dessin. Prenons $\mathrm{I} = [0, 1]$, et traçons le graphe d'une fonction dérivable, telle que la pente de la tangente reste petite : d'après le lemme 6.17, elle est lipschitzienne. On fait en sorte que la fonction prenne ses valeurs dans $[0, 1]$, bien sûr. Ajoutons au dessin un point u_0 et son image $f(u_0)$, ainsi que la diagonale (l'ensemble des points $\{(x, x) \mid x \in [0, 1]\}$).

Pour dessiner les points suivants de la suite, on doit reporter $f(u_0)$ sur l'axe des abscisses. Pour cela, nous devons en fait prendre le symétrique du point $(0, f(u_0))$, qui se trouve déjà sur la figure sur l'axe des ordonnées, par rapport à la diagonale.

Nous allons poursuivre en conservant uniquement les segments indiqués en traits pleins. La recette est simple : on part d'un point du graphe, on rejoint la diagonale, puis on vire à angle droit en direction du graphe. On obtient une figure en forme de spirale, qui montre bien la convergence de la suite vers le point à l'intersection du graphe et de la diagonale, c'est-à-dire le point fixe.

À titre d'exercice, tentez l'expérience suivante : essayez de faire le même dessin mais avec un graphe tel que la dérivée au « point fixe » est −2 ou −3. Que se passe-t-il ?

Exemple 6.19 (Calcul numérique de racines carrées) – Il arrive parfois que l'intérêt du théorème ne soit pas dans l'existence et l'unicité du point fixe, qui peuvent être évidentes pour d'autres raisons, mais dans la méthode de calcul de ce point fixe qui est proposée, à l'aide de la suite (u_n).

Voici un exemple. Considérons la fonction f définie pour $x > 0$ par $f(x) = \frac{1}{2}(x + \frac{\lambda}{x})$, où λ est un réel positif. L'équation $f(x_0) = x_0$ est équivalente à $x_0^2 = \lambda$, qui pour $x_0 > 0$ possède la solution unique $x_0 = \sqrt{\lambda}$. Oui, mais comment évaluer numériquement $\sqrt{\lambda}$? Il existe plusieurs façons de procéder évidemment, mais le théorème du point fixe en fournit déjà une qui est efficace.

Pour se ramener au cadre du théorème, prenons $a = \sqrt{\lambda}$ et $b > \sqrt{\lambda}$ quelconque, et posons $\mathrm{I} = [a, b]$. La dérivée est donnée par $f'(x) = \frac{1}{2}(1 - \frac{\lambda}{x^2})$, et on vérifie alors que $0 \leq f'(x) \leq \frac{1}{2}$ pour $x \in \mathrm{I}$. On en déduit que f est $\frac{1}{2}$-lipschitzienne sur I. De plus, f est croissante sur cet intervalle ; comme $f(a) = a$, et puisqu'on voit de suite que $f(b) \leq b$, l'intervalle I vérifie $f(\mathrm{I}) \subset \mathrm{I}$.

À partir de maintenant, on voit f comme une fonction $\mathrm{I} \to \mathrm{I}$ à laquelle le théorème s'applique. En prenant par exemple $u_0 = b$, nous savons maintenant que la suite définie par récurrence par $u_{n+1} = f(u_n) = \frac{1}{2}(u_n + \frac{\lambda}{u_n})$ converge vers $\sqrt{\lambda}$.

Essayons pour $\lambda = 2$. En prenant $b = 2$, on en déduit que la suite vérifiant $u_0 = 2$ et $u_{n+1} = \frac{1}{2}(u_n + \frac{2}{u_n})$ converge vers $\sqrt{2}$. Les premiers termes sont

$$u_0 = 2, \qquad u_1 = \frac{3}{2} = 1{,}5, \qquad u_2 = \frac{17}{12} = 1{,}41666\ldots$$

puis

$$u_3 = \frac{577}{408} = 1{,}41421568\ldots, u_4 = \frac{665857}{470832} = 1{,}41421356\ldots$$

De plus, on peut calculer la précision du résultat, ce qui est un grand mérite de la méthode. En effet, on a

$$|u_{n+1} - x_0| = |f(u_n) - f(x_0)| \leq k|u_n - x_0|,$$

et on en déduit (un peu comme dans la démonstration du théorème) que

$$|u_n - x_0| \leq k^n|u_0 - x_0|.$$

Dans notre exemple, on a pris $k = \frac{1}{2}$ et donc on sait que

$$|u_n - \sqrt{2}| \leq \frac{1}{2^n}|2 - \sqrt{2}| \leq \frac{1}{2^n}.$$

Comme $2^{14} > 10000$, on en déduit que l'écart entre u_{14} et $\sqrt{2}$ est inférieur à $0{,}0001$, par exemple ; puisque u_{14} commence par $1{,}414$, on en déduit que $\sqrt{2}$ commence également par $1{,}414$. (Et sachant ceci, on en déduit *a posteriori* que u_3 commence déjà par les 3 bons chiffres après la virgule ; mais il fallait bien le démontrer… En réalité on peut montrer que dans ce cas particulier, le calcul est une application de la « méthode de Newton » et que celle-ci converge très vite. Les 5 premiers chiffres de u_3 sont en fait corrects !)

Proposition 6.20 – *Soit $f : \mathrm{I} \to \mathrm{J}$ une bijection, et soit $f^{-1} : \mathrm{J} \to \mathrm{I}$ sa réciproque. On suppose que f est continue, et qu'elle est dérivable au point $x_0 \in \mathrm{I}$, avec $f'(x_0) \neq 0$. Alors en notant $y_0 = f(x_0)$, la fonction f^{-1} est dérivable au point y_0 et*

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

Démonstration. Le taux d'accroissement pour f^{-1} au point y_0 est la fonction T_{y_0} définie par

$$\mathrm{T}_{y_0}(y) = \frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0},$$

pour $y \neq y_0$. On a donc

$$\mathrm{T}_{y_0}(f(x)) = \frac{f^{-1}(f(x)) - f^{-1}(f(x_0))}{f(x) - f(x_0)} = \frac{x - x_0}{f(x) - f(x_0)},$$

et cette expression a un sens pour $x \neq x_0$ puisqu'alors $f(x) \neq f(x_0)$. Par dérivabilité de f, on en déduit que $\mathrm{T}_{y_0}(f(x)) \to \frac{1}{f'(x_0)}$ lorsque $x \to x_0$.

Comme la fonction f^{-1} est continue par le théorème 4.24, on a $f^{-1}(y) \to f^{-1}(y_0) = x_0$ lorsque $y \to y_0$. Par composition,

$$\mathrm{T}_{y_0}(y) = \mathrm{T}_{y_0}\left[f(f^{-1}(y))\right] \longrightarrow \frac{1}{f'(x_0)}$$

lorsque $y \to y_0$, ce qui signifie bien que f^{-1} est dérivable en y_0, et que le nombre dérivé est celui annoncé. □

En combinant plusieurs résultats déjà obtenus, on en arrive au théorème suivant, qui est facile à mémoriser.

Théorème 6.21 – *Soit f une fonction continument dérivable sur l'intervalle I, telle que $f'(x) \neq 0$ pour tout $x \in \mathrm{I}$.*

Alors f réalise une bijection $\mathrm{I} \to \mathrm{J}$. Sa réciproque f^{-1} est également continument dérivable, et

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))},$$

pour tout $y \in \mathrm{J}$.

Enfin, si I est ouvert, alors J l'est aussi.

Notez que l'on dit d'une fonction f qu'elle est *continument dérivable* pour indiquer qu'elle est dérivable et que f' est continue. On parle aussi parfois de « fonction de classe C^1 ».

Avant 1990, l'orthographe recommandée était « continûment ».

Démonstration. Puisque la fonction f' est continue et ne s'annule pas, elle ne peut pas changer de signe en vertu du théorème des valeurs intermédiaires : ainsi, ou bien $f'(x) > 0$ pour tout $x \in \mathrm{I}$ et f est strictement croissante, ou bien $f'(x) < 0$ et f est strictement décroissante. Dans tous les cas, f est strictement monotone, et donc injective. En notant $\mathrm{J} = f(\mathrm{I})$, qui est un intervalle encore d'après le théorème des valeurs intermédiaires, on a donc établi que $f : \mathrm{I} \to \mathrm{J}$ est une bijection.

La proposition précédente montre que f^{-1} est dérivable en tout point de J, et montre également la formule pour $(f^{-1})'$. Le lemme 6.3 montre que f^{-1} est continue (le théorème 4.24 aussi), et on en conclut que $(f^{-1})' = \frac{1}{f' \circ f^{-1}}$ est elle-même continue.

Enfin, une fonction strictement monotone ne peut pas avoir de minimum ni de maximum sur un intervalle ouvert (vérifiez-le), donc J doit être ouvert si I est ouvert. □

Exemple 6.22 – On peut retrouver certains des résultats de la proposition 6.11. Par exemple, sachant que la fonction exponentielle est dérivable, et que c'est même sa propre dérivée, on peut calculer la dérivée de sa réciproque, le logarithme, à partir du théorème :

$$\ln'(x) = \frac{1}{\exp'(\ln(x))} = \frac{1}{\exp(\ln(x))} = \frac{1}{x}.$$

Autre exemple, sachant que la fonction tangente est dérivable et que

$$\tan'(x) = 1 + (\tan(x))^2 > 0,$$

on en déduit que sa réciproque arctangente est dérivable et que

$$\arctan'(x) = \frac{1}{\tan'(\arctan(x))} = \frac{1}{1 + \tan(\arctan(x))^2} = \frac{1}{1 + x^2}.$$

Sur le même modèle, vous pourrez traiter arcsinus et arccosinus.

Enfin, prenons $f(x) = x^n$ (pour $n \geq 1$), de sorte que $f'(x) = nx^{n-1}$. Cette dérivée s'annule en 0 (pour $n \geq 2$), donc pour appliquer le théorème il faut se restreindre à $]0, +\infty[$. On en déduit que la réciproque, c'est-à-dire la fonction $f^{-1}(x) = \sqrt[n]{x}$, est dérivable pour $x > 0$ et que

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))} = \frac{1}{n(f^{-1}(x))^{n-1}} = \frac{1}{n(\sqrt[n]{x})^{n-1}} = \frac{1}{n}(\sqrt[n]{x})^{1-n}.$$

En 0, cette fonction n'est pas dérivable, voir l'exemple 6.4.

Soit $I \subset \mathbb{R}$, et soit $f : I \to \mathbb{R}^n$, qui est donc de la forme

$$t \mapsto f(t) = (f_1(t), f_2(t), \ldots, f_n(t)).$$

Pour définir la dérivée d'une telle fonction, deux alternatives peuvent venir à l'esprit ; et la proposition suivante affirme qu'elles coïncident.

Proposition 6.23 – *Avec les notations ci-dessus, les deux propriétés suivantes sont équivalentes :*

1. *chaque fonction f_i est dérivable au point t_0 ;*
2. *la fonction*

$$t \mapsto \frac{f(t) - f(t_0)}{t - t_0}$$

possède une limite lorsque $t \to t_0$.

Dans ce cas, la limite en question est

$$(f_1'(t_0), \ldots, f_n'(t_0)),$$

que l'on appelle le vecteur dérivé de f en t_0 ; on le note $f'(t_0)$.

Rappelons que la notion de limite de fonction à valeurs vectorielles a été donnée dans la définition 4.26. Il faut bien comprendre que lorsque l'on forme

$$\frac{f(t) - f(t_0)}{t - t_0},$$

le numérateur est une différence de vecteurs (ou de matrices-colonnes, si l'on veut), alors que le dénominateur est un scalaire. C'est-à-dire que

$$\frac{f(t) - f(t_0)}{t - t_0} = \frac{1}{t - t_0} \begin{pmatrix} f_1(t) - f_1(t_0) \\ \vdots \\ f_n(t) - f_n(t_0) \end{pmatrix}.$$

Ayant réalisé ceci, la démonstration est facile.

Exemple 6.24 – Une fonction de la forme $\gamma : I \to \mathbb{R}^2$, donc de la forme

$$t \mapsto \gamma(t) = (x(t), y(t)),$$

est appelée une *courbe*. La dérivée, lorsqu'elle existe, est

$$t \mapsto \gamma'(t) = (x'(t), y'(t)),$$

et $\gamma'(t)$ est appelé le *vecteur-vitesse* à l'instant t : en effet on peut penser à γ comme à un point qui se déplace dans le plan. La *vitesse* à l'instant t est

$$\|\gamma'(t)\| = \sqrt{x'(t)^2 + y'(t)^2}.$$

Par abus de langage, l'image de γ, c'est-à-dire l'ensemble

$$\gamma(I) = \{\gamma(t) \mid t \in I\},$$

est parfois également appelé une courbe. Au point $\gamma(t)$, on peut tracer la droite de vecteur directeur $\gamma'(t)$, qui est appelée la tangente à la courbe à l'instant t. Nous allons étudier quelques courbes dans les exercices. Ça se fait simplement en étudiant séparément les fonctions $t \mapsto x(t)$ et $t \mapsto y(t)$.

voir l'exercice 5524

Les propriétés usuelles des dérivées restent vraies pour les fonctions à valeurs vectorielles, par exemple il est clair que $(f+g)' = f' + g'$. Il n'y a pas de formule pour le produit, puisque le produit de deux vecteurs n'est pas en général défini.

Par contre, on peut considérer les fonctions dont les valeurs sont des *matrices*, c'est-à-dire les fonctions du type $I \to M_{n,m}(\mathbb{R})$. On peut identifier $M_{n,m}(\mathbb{R})$ avec $\mathbb{R}^{nm}$, et ce qui précède s'applique. On a alors sans surprise :

Lemme 6.25 – *Si $f : I \to M_{n,m}(\mathbb{R})$ et $g : I \to M_{m,\ell}(\mathbb{R})$ sont dérivables, alors $fg : I \to M_{n,\ell}(\mathbb{R})$ est dérivable et*

$$(fg)'(t) = f'(t)g(t) + f(t)g'(t).$$

La démonstration peut se faire directement par le calcul, en se basant sur la formule pour les fonctions à valeurs réelles (voir la proposition 6.5).

Par contre, il faut faire attention à une chose : le théorème des accroissements finis concerne strictement les fonctions à valeurs réelles, et n'a pas d'équivalent pour les fonctions vectorielles. Il reste néanmoins vrai que si une fonction dérivable f sur un intervalle I vérifie $f'(t) = 0$ pour tout $t \in I$, alors f est constante : en effet il suffit de le vérifier pour chaque composante de f.

Enfin, concluons en indiquant que ces dernières remarques sur les fonctions à valeurs vectorielles ou matricielles restent vraies en remplaçant $\mathbb{R}$ par $\mathbb{C}$.

Chapitre 7

Formules de Taylor

Soit f une fonction définie sur un intervalle I contenant 0. Nous allons examiner les conditions de continuité et de dérivabilité en 0 sous un angle un peu nouveau.

La fonction f est continue en 0 si et seulement si $\lim f(x) = f(0)$ lorsque $x \to 0$. Dans ce cas, on peut écrire, même si ça paraît artificiel pour l'instant, que

$$f(x) = f(0) + \varepsilon(x),$$

avec $\varepsilon(x) = f(x) - f(0)$; on observe alors que $\varepsilon(x) \to 0$ lorsque $x \to 0$. En d'autres termes, la fonction f se rapproche de la valeur (constante) $f(0)$ lorsque x s'approche de 0. On ne sait pas à quelle vitesse cette approche se fait.

De la même manière, le lemme 6.7 nous dit que si f est dérivable en 0, alors

$$f(x) = f(0) + f'(0)x + x\varepsilon(x),$$

où là encore $\varepsilon(x) \to 0$ lorsque $x \to 0$. Écrivons $\mathrm{P}(x) = f(0) + f'(0)x$; c'est un polynôme en x, de degré 1. Comme nous le faisions remarquer après le lemme 6.7, la différence $f(x) - \mathrm{P}(x) = x\varepsilon(x)$ est le produit de deux fonctions qui tendent vers 0 avec x, et on peut donc considérer que P est une approximation assez bonne de f.

Géométriquement, le graphe de P est la droite passant par $(0, f(0))$ et dont le coefficient directeur est $f'(0)$; c'est la droite tangente au graphe de f. On en sait donc un peu plus sur la façon dont f approche la valeur $f(0)$ lorsque $x \to 0$.

Le but de ce chapitre est de montrer que l'on peut continuer dans cette voie : on peut trouver, en supposant que l'on peut dériver f au moins n fois, un polynôme $\mathrm{P}_n(x)$ de degré $\leq n$ tel que la différence $f(x) - \mathrm{P}_n(x)$ tende vers 0 encore plus vite.

C'est utile dans de nombreux calculs : nous allons être capables de calculer des limites qui restent inaccessibles par les autres méthodes classiques. En fait, presque n'importe quel calcul de limite va se ramener à une limite de polynômes !

Il y a un certain nombre de formules, dites « de Taylor », qui ont toutes pour objectif l'approximation d'une fonction par un polynôme, comme annoncé dans l'introduction. Nous en verrons deux, il en existe encore d'autres.

Théorème 7.1 (Taylor-Lagrange) – *Soit f une fonction dérivable n fois sur un intervalle I contenant 0. Alors $\forall x \in \mathrm{I}$ il existe $\theta \in]0,1[$ tel que :*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \frac{f^{(n)}(\theta x)}{n!}x^n.$$

(*Attention* : θ dépend de x et de n.)

Démonstration. On va faire la démonstration pour $x > 0$, pour simplifier.

On peut toujours trouver un réel A (qui dépend de x !) tel que

$$f(x) = f(0) + f'(0)x + \cdots + \frac{\mathrm{A}}{n!}x^n.$$

(En effet il suffit de poser $\mathrm{A} = (f(x) - (f(0) + \cdots))/(x^n/n!)$.) On veut montrer que $\mathrm{A} = f^{(n)}(\theta x)$ pour un certain $0 < \theta < 1$.

Sur l'intervalle $[0, x]$, on définit

$$\mathrm{F} : t \mapsto f(x) - \left[f(t) + f'(t)(x-t) + \frac{f''(t)}{2}(x-t)^2 + \cdots + \frac{f^{(n-1)}(t)}{(n-1)!}(x-t)^{n-1}\right] - \frac{\mathrm{A}}{n!}(x-t)^n.$$

Cette fonction F est dérivable, on a $\mathrm{F}(0) = 0$ par choix de A, et $\mathrm{F}(x) = 0$. Le théorème des accroissements finis donne l'existence de $c \in]0, x[$ tel que

$$\mathrm{F}'(c) = \frac{\mathrm{F}(x) - \mathrm{F}(0)}{x - 0} = 0.$$

En posant $\theta = \frac{c}{x}$, on a bien $0 < \theta < 1$ et $c = \theta x$.

Calculons maintenant $\mathrm{F}'(t)$ (les détails vous sont confiés en exercice) :

$$\mathrm{F}'(t) = \frac{\mathrm{A}}{(n-1)!}(x-t)^{n-1} - \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} = \left(\mathrm{A} - f^{(n)}(t)\right)\frac{(x-t)^{n-1}}{(n-1)!}.$$

Donc l'équation $\mathrm{F}'(\theta x) = 0$ que nous avons obtenue donne $\mathrm{A} = f^{(n)}(\theta x)$, comme on le souhaitait. □

Pour $n = 1$, noter que cette formule redonne exactement le théorème des accroissements finis (un tout petit peu reformulé).

Exemple 7.2 – Prenons la fonction $f(x) = e^x$. On a $f' = f$, et donc aussi $f'' = f$ et par récurrence $f^{(n)} = f$ pour tout n. En particulier $f^{(n)}(0) = 1$. La formule de Taylor-Lagrange donne donc, pour tout $x \in \mathbb{R}$ et tout entier n , l'existence d'un nombre $0 < \theta < 1$ (qui dépend de x et de n) tel que

$$f(x) = e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^{n-1}}{(n-1)!} + \frac{e^{\theta x}}{n!}x^n.$$

En guise d'application, essayons de fixer x et de faire tendre n vers l'infini. Si on observe que $|e^{\theta x}| \le e^{|\theta x|} \le e^{|x|}$, on en tire

$$\left| e^x - \sum_{k=0}^{n-1}\frac{x^k}{k!} \right| \le e^{|x|}\frac{|x|^n}{n!}.$$

Puisque $\frac{|x|^n}{n!} \to 0$ lorsque $n \to +\infty$ (toujours avec x fixé), on en déduit

$$e^x = \lim_{n\to+\infty}\sum_{k=0}^{n-1}\frac{x^k}{k!} = \sum_{k=0}^{+\infty}\frac{x^k}{k!}.$$

On retrouve la définition de l'exponentielle telle que nous l'avions donnée (définition B.1, pour ceux qui ont pris connaissance de l'appendice B). De plus, nous avons simplement utilisé le fait que $f' = f$ et $f(0) = 1$, donc nous retrouvons la partie « unicité » du lemme B.4.

Exemple 7.3 – Prenons maintenant une fonction pour laquelle nous ne connaissons pas encore de développement en série : par exemple $f(x) = \ln(1 + x)$, définie sur $]-1, +\infty[$. On a

$$f'(x) = \frac{1}{1+x} \text{ donc } f''(x) = -\frac{1}{(1+x)^2} \text{ et } f^{(3)}(x) = \frac{2}{(1+x)^3}.$$

On peut montrer que $f^{(n)}(x) = (-1)^{n-1}(n-1)!(1+x)^{-n}$; en fait en dérivant cette formule on obtient tout de suite la forme de $f^{(n+1)}(x)$, d'où le résultat par récurrence. En particulier, on a $f^{(n)}(0) = (-1)^{n-1}(n-1)!$.

La formule de Taylor-Lagrange donne l'existence pour tout x et tout n d'un nombre $0 < \theta < 1$ tel que

$$f(x) = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^n\frac{x^{n-1}}{n-1} + \frac{(-1)^{n+1}x^n}{n(1+\theta x)^n}.$$

Là encore, fixons x et faisons tendre n vers $+\infty$, pour voir. On va se restreindre à $x \in [0, 1]$, et en particulier $|x^n| \le 1$. Dans ce cas on a

$$\left| \frac{(-1)^{n-1}x^n}{n(1+\theta x)^n} \right| \le \frac{1}{n} \xrightarrow[n\to\infty]{} 0.$$

Ce qui signifie pour tout $x \in [0, 1]$ que :

$$\ln(1+x) = \lim_{n\to+\infty}\sum_{k=1}^{n-1}(-1)^{k-1}\frac{x^k}{k} = \sum_{k=1}^{+\infty}(-1)^{k-1}\frac{x^k}{k}.$$

(Il est possible de montrer que ceci est vrai pour tout x dans l'intervalle $]-1, 1]$, mais nous ne le ferons pas ici.) C'est cette formule qui est utilisée par les calculatrices pour calculer un logarithme ! Notons pour $x = 1$ que l'on a

$$\ln(2) = \sum_{k=1}^{+\infty}\frac{(-1)^{k-1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

Nous avions annoncé ce résultat dans l'exemple 3.23.

Dans les exemples ci-dessus, nous avons fixé x et étudié le « reste », c'est-à-dire le terme en x^n ; c'est typique de l'utilisation de Taylor-Lagrange. Voici maintenant la formule de Taylor-Young, qui va être utilisée lorsque l'on veut faire tendre x vers 0 pour calculer une limite.

Théorème 7.4 (Taylor-Young) – *Soit f une fonction dérivable n fois sur un intervalle I contenant 0. Alors on peut écrire*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + x^n\varepsilon(x)$$

où $\varepsilon(x) \to 0$ quand $x \to 0$.

Démonstration. La démonstration complète sera donnée plus loin (voir le théorème 8.36). Ici nous donnons une démonstration avec une toute petite restriction : on va supposer que, en plus des hypothèses ci-dessus, la fonction $f^{(n)}$ est continue en 0. C'est par exemple le cas si f est dérivable $n+1$ fois, et en pratique dans tous nos exemples nous serons dans cette situation.

On a alors $f^{(n)}(x) \to f^{(n)}(0)$; en d'autres termes, si on pose $h(x) = f^{(n)}(x) - f^{(n)}(0)$, on peut écrire $f^{(n)}(x) = f^{(n)}(0) + h(x)$ avec $h(x) \to 0$. La formule de Taylor-Lagrange donne alors :

$$f(x) = f(0) + \cdots + \frac{f^{(n-1)}(0)}{(n-1)!}x^{n-1} + \frac{f^{(n)}(0)}{n!}x^n + \frac{x^n h(\theta_x x)}{n!}.$$

On rappelle encore que $\theta = \theta_x$ dépend de x. Posons alors $\varepsilon(x) = h(\theta_x x)/n!$. Comme $0 < \theta_x < 1$, on a $\varepsilon(x) \to 0$ lorsque $x \to 0$, ce qui conclut la démonstration. □

Exemple 7.5 – Avant de connaître les formules de Taylor, calculer une limite relève souvent de l'astuce. Par exemple essayons de calculer la limite de $\frac{\cos(x)-1}{x}$ lorsque $x \to 0$. L'astuce consiste à remarque que, si $f(x) = \cos(x)$, alors la quantité étudiée est le taux d'accroissement de f :

$$\frac{\cos(x)-1}{x} = \frac{f(x)-f(0)}{x-0} \to f'(0) = 0.$$

Si maintenant nous essayons de calculer la limite de $\frac{\cos(x)-1}{x^2}$ en 0, la même astuce ne donne rien. Pour parvenir à faire le calcul, écrivons la formule de Taylor-Young pour la fonction f définie par $f(x) = \cos(x)$ et pour $n = 2$ (on dit souvent « Taylor-Young à l'ordre 2 »). On a $f(0) = 1$, $f'(0) = 0$ et $f''(0) = -1$. Par conséquent :

$$f(x) = \cos(x) = 1 - \frac{x^2}{2} + x^2\varepsilon(x)$$

avec $\varepsilon(x) \to 0$. Et donc :

$$\frac{\cos(x)-1}{x^2} = -\frac{1}{2} + \varepsilon(x) \longrightarrow -\frac{1}{2}.$$

Ainsi, la formule de Taylor nous a permis de « remplacer » la fonction cos par le polynôme $1 - \frac{x^2}{2}$ dans le calcul de la limite.

Au passage, il est très facile d'écrire la formule de Taylor-Young pour la fonction cos à n'importe quel ordre, puisque les dérivées successives sont :

$$\cos' = -\sin, \quad \cos'' = -\cos, \quad \cos^{(3)} = \sin \quad \text{et} \quad \cos^{(4)} = \cos.$$

Les nombres $\cos^{(n)}(0)$, lorsque n augmente, sont donc

$$1, 0, -1, 0, \quad 1, 0, -1, 0, \quad 1, 0, -1, 0, \ldots,$$

la séquence $1, 0, -1, 0$ se répétant sans cesse. En particulier les termes « impairs » dans la formule de Taylor-Young sont nuls, et la formule à l'ordre $2n+1$ est

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n\frac{x^{2n}}{(2n)!} + x^{2n+1}\varepsilon(x),$$

où $\varepsilon(x) \to 0$ lorsque $x \to 0$.

Exemple 7.6 – On souhaite calculer la limite en 0 de $\frac{\sin(x)-x}{x^2}$. Écrivons d'abord la formule de Taylor-Young pour la fonction sinus : les dérivées successives sont

$$\sin' = \cos, \quad \sin'' = -\sin, \quad \sin^{(3)} = -\cos \quad \text{et} \quad \sin^{(4)} = \sin.$$

Les nombres $\sin^{(n)}(0)$, lorsque n augmente, sont donc

$$0, 1, 0, -1, \quad 0, 1, 0, -1, \quad 0, 1, 0, -1, \ldots.$$

En particulier les termes « pairs » dans la formule de Taylor-Young sont nuls, et la formule à l'ordre $2n$ est

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + x^{2n}\varepsilon(x),$$

où $\varepsilon(x) \to 0$ lorsque $x \to 0$. Pour notre limite, prenons l'ordre 4 :

$$\sin(x) = x - \frac{x^3}{6} + x^4\varepsilon(x),$$

où, vous l'aurez deviné, on a $\varepsilon(x) \to 0$. Ainsi

$$\frac{\sin(x)-x}{x^2} = -\frac{x}{6} + x^2\varepsilon(x) \longrightarrow 0.$$

Exemple 7.7 – Un dernier. Essayons de calculer la limite de

$$\frac{\sqrt{1+2x} - \sqrt[3]{1+3x}}{x^2},$$

lorsque x tend vers 0. Introduisons la notation $f_\alpha(t) = (1+t)^\alpha$, pour $\alpha > 0$, et calculons la formule de Taylor-Young pour f_α. On a

$$f_\alpha'(t) = \alpha(1+t)^{\alpha-1}, \qquad f_\alpha''(t) = \alpha(\alpha-1)(1+t)^{\alpha-2},$$

et par récurrence on a facilement

$$f_\alpha^{(n)}(t) = \alpha(\alpha-2)(\alpha-2)\cdots(\alpha-n+1)(1+t)^{\alpha-n}.$$

Le coefficient qui apparaît dans Taylor-Young est donc

$$\frac{f_\alpha^{(n)}(0)}{n!} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-n+1)}{n!},$$

et ce nombre est souvent noté $\binom{\alpha}{n}$, ce qui est cohérent avec la notation lorsque α est un entier. On a donc

$$(1+t)^\alpha = 1 + \alpha t + \frac{\alpha(\alpha-1)}{2}t^2 + \cdots + \binom{\alpha}{n}t^n + t^n\varepsilon(t),$$

avec $\varepsilon(t) \to 0$. Pour $\alpha = \frac{1}{2}$ et $n = 2$, on obtient

$$\sqrt{1+t} = (1+t)^{\frac{1}{2}} = 1 + \frac{1}{2}t - \frac{t^2}{8} + t^2\varepsilon_1(t).$$

Pour $t = 2x$ ceci donne

$$\sqrt{1+2x} = 1 + x - \frac{x^2}{2} + 4x^2\varepsilon_1(2x).$$

Vous montrerez qu'en prenant $\alpha = \frac{1}{3}$ on en arrive à

$$\sqrt[3]{1+3x} = 1 + x - x^2 + 9x^2\varepsilon_2(3x).$$

Finalement l'expression dont on cherche la limite est de la forme

$$\frac{\frac{1}{2}x^2 + x^2h(x)}{x^2} = \frac{1}{2} + h(x),$$

où $h(x)$ est une certaine expression qui tend vers 0 avec x. La limite vaut $\frac{1}{2}$.

Si vous avez trouvé ce dernier calcul un peu compliqué, alors vous conviendrez qu'on aurait besoin de notations plus simples, et de quelques conseils pratiques.

Les expressions telles que $4x^2\varepsilon_1(2x)$ ci-dessus sont rapidement pénibles à manier. Donner des noms différents aux fonctions qui tendent vers 0 qui apparaissent (ε_1, ε_2, ...) devient vite compliqué, et on se demande s'il est vraiment utile de baptiser toutes ces fonctions. On ne peut pourtant pas outes les nommer de la même manière.

Pour résoudre ce problème, on introduit la *notation de Landau*, qui en toute rigueur est un peu ambigüe, mais en pratique économise bien des efforts. Elle fonctionne de la manière suivante : tout d'abord on écrit

$$o(1)$$

qui se prononce « petit o de 1 », pour désigner une fonction anonyme qui tend vers 0. On ne dit pas « quand qui tend vers quoi », c'est pourquoi la notation est ambigüe, mais c'est le contexte qui rend les choses claires.

Ensuite, étant donnée une fonction ϕ, qui en pratique sera très souvent de la forme $\phi(x) = x^n$, on utilise le raccourci

$$o(\phi(x)) = \phi(x)\,o(1).$$

Par exemple $o(x^n)$ désigne une expression de la forme $x^n\varepsilon(x)$ avec $\varepsilon(x) \to 0$ (et ces expressions sont beaucoup intervenues dans le début de ce chapitre !). Ainsi on peut énoncer la conclusion du théorème de Taylor-Young sous la forme

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + o(x^n).$$

On peut penser à $o(\phi(x))$ comme à « quelque chose de négligeable devant $\phi(x)$ ».

Avant de voir cette notation à l'oeuvre dans un calcul, une petite définition :

Définition 7.8 – On dit que f a *un développement limité* à l'ordre n au voisinage de 0 s'il existe $a_0, \ldots, a_n \in \mathbb{R}$ tels que

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n + o(x^n).$$

Le théorème de Taylor-Young affirme donc que si f est dérivable n fois, alors elle possède un développement limité à l'ordre n, et de plus $a_k = \frac{f^{(k)}(0)}{k!}$. Mais utiliser le théorème n'est pas toujours la meilleure façon de trouver un développement limité – en fait, presque jamais.

Exemple 7.9 – En écrivant $(1-x)(1+x+x^2+\cdots+x^n) = 1-x^{n+1}$, on tire

$$\begin{aligned}\frac{1}{1-x} &= 1 + x + x^2 + \cdots + x^n + \frac{x^{n+1}}{1-x}\\ &= 1 + x + x^2 + \cdots + x^n + o(x^n).\end{aligned}$$

Ici on a utilisé le fait que $\frac{x}{1-x} \to 0$ lorsque $x \to 0$, ce qu'on résume en écrivant $\frac{x}{1-x} = o(1)$; et donc $\frac{x^{n+1}}{1-x} = o(x^n)$.

Bien entendu on peut obtenir ce développement grâce à Taylor-Young, ou encore en utilisant celui de $(1+t)^\alpha$ comme dans l'exemple 7.7 pour $\alpha = -1$ et $t = -x$. Mais le plus simple pour le retrouver reste le calcul ci-dessus.

Exemple 7.10 – Cherchons un développement limité de la fonction $x \mapsto (e^x-1)(\sin(x)-x)$ en 0 à l'ordre 4. On écrit

$$e^x - 1 = x + o(x) \quad \text{et} \quad \sin(x) - x = -\frac{x^3}{6} + o(x^3) ;$$

en effet nous connaissons ces développements par cœur depuis les exemples 7.2 et 7.6.

En multipliant il vient

$$(e^x-1)(\sin(x)-x) = -\frac{x^4}{6} + \left[-\frac{x^3}{6}o(x) + xo(x^3) + o(x)o(x^3)\right].$$

Maintenant nous faisons une série de petites simplifications, qu'il va falloir s'habituer à faire de tête (c'est très facile). Tout d'abord

$$-\frac{x^3}{6}o(x) = -\frac{x^4}{6}o(1) = x^4o(1) = o(x^4).$$

Pour la deuxième égalité, on utilise le fait que $-\frac{1}{6}o(1) = o(1)$, ce qui signifie seulement que $-\frac{1}{6}o(1)$ tend vers 0 avec x.

Deux autres petits calculs donnent de la même manière $xo(x^3) = o(x^4)$ et $o(x)o(x^3) = o(x^4)$. Enfin la somme des termes dans le crochet est

$$o(x^4) + o(x^4) + o(x^4) = x^4(o(1) + o(1) + o(1)) = x^4o(1) = o(x^4).$$

Ici on utilise le fait que $o(1) + o(1) = o(1)$ (ce qui surprend la première fois !). Finalement

$$(e^x-1)(\sin(x)-x) = -\frac{x^4}{6} + o(x^4).$$

Proposition 7.11 – *Si l'on peut écrire :*

$$\begin{aligned}f(x) &= a_0 + a_1x + \cdots + a_nx^n + o(x^n)\\ &= b_0 + b_1x + \cdots + b_nx^n + o(x^n)\end{aligned}$$

alors $a_i = b_i$. En d'autres termes, lorsqu'un développement limité existe, il est unique.

Démonstration. Par récurrence sur n. Pour $n = 0$, on prend la limite quand x tend vers 0, et on obtient $a_0 = b_0$.

Si l'unicité a été prouvée pour $n-1$ et que l'on a un développement limité à l'ordre n, on écrit $a_nx^n + o(x^n) = o(x^{n-1})$ et de même pour b_n, et on obtient

$$\begin{aligned}f(x) &= a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + o(x^{n-1})\\ &= b_0 + b_1x + \cdots + b_{n-1}x^{n-1} + o(x^{n-1}).\end{aligned}$$

Par récurrence on a $a_i = b_i$ pour $0 \le i \le n-1$.

On peut donc simplifier l'égalité ci-dessus, et il reste $a_nx^n + o(x^n) = b_nx^n + o(x^n)$. On divise par x^n et on prend la limite en 0 : il vient $a_n = b_n$. □

Exemple 7.12 – Cette proposition est utile, bien sûr, lorsque nous avons deux façons de trouver un développement limité. Par exemple, considérons la question suivante : soit $f(x) = \frac{1}{1+2x^3}$; combien vaut $f^{(6)}(0)$? On peut bien sûr répondre à cette question en dérivant 6 fois... mais c'est très long. Procédons autrement.

D'après Taylor-Young, nous savons que cette fonction possède un développement limité à tous les ordres, et le terme en x^k est précisément $\frac{f^{(k)}(0)}{k!}x^k$. Mais nous savons aussi que

$$\frac{1}{1-u} = 1 + u + u^2 + o(u^2),$$

depuis l'exemple 7.9, donc en prenant $u = -2x^3$:

$$\frac{1}{1+2x^3} = 1 - 2x^3 + 4x^6 + o(x^6).$$

Détaillons un peu ce qui vient de se passer avec le reste. Le terme $o(u^2)$ s'écrit donc $u^2\varepsilon(u)$ avec $\varepsilon(u) \to 0$ lorsque $u \to 0$. Lorsque l'on fait $u = -2x^3$, ce terme devient $4x^6\varepsilon(-2x^3)$, et cette expression est bien de la forme $x^6o(1) = o(x^6)$. Là encore il faut faire ça de tête, avec l'habitude.

D'après la proposition, on peut comparer ce développement limité avec celui donné par Taylor-Young, et en particulier pour les termes en x^6 la comparaison donne

$$\frac{f^{(6)}(0)}{6!} = 4 \quad \text{donc} \quad f^{(6)}(0) = 4 \times 6! = 2880.$$

Calculer les développements limités nécessite de l'entraînement, et nous allons lister quelques techniques à connaître. Évidemment la première méthode est d'appliquer le théorème de Taylor-Young, mais il est très rare que ce soit le meilleur choix. C'est bien sûr grâce à ce théorème que nous avons obtenu les développements des fonctions usuelles (exponentielle, sinus, cosinus…), mais ceux-ci sont à savoir par cœur absolument, de sorte qu'à partir de maintenant il sera exceptionnel d'appliquer directement Taylor-Young. (À la fin du chapitre nous résumons les choses à mémoriser.)

EXEMPLE 7.13 (COMPOSITION) – Essayons de trouver un développement limité de $\frac{1}{\cos(x)}$ à l'ordre 4 en 0. On utilise le fait que :

$$\frac{1}{1+u} = 1 - u + u^2 + o(u^2).$$

Ensuite on écrit $\cos(x) = 1 + u(x)$ avec $u(x) = -\frac{x^2}{2} + \frac{x^4}{4!} + o(x^4)$ (on a retenu le calcul de l'exemple 7.5 par cœur).

En combinant les résultats, on obtient

$$\begin{aligned}\frac{1}{\cos(x)} = \frac{1}{1+u(x)} &= 1 - u(x) + u(x)^2 + o(u(x)^2)\\ &= 1 + \frac{x^2}{2} + \frac{5x^4}{24} + o(x^4).\end{aligned}$$

On a utilisé au passage $o(u(x)^2) = o(x^4)$. D'une manière générale le petit résultat suivant est à retenir : si $u(x)$ possède un développement limité qui commence par un terme en x^m, alors $o(u(x)^n) = o(x^{nm})$.

EXEMPLE 7.14 (INTÉGRATION) – Le principe est le suivant. Soit f une fonction dérivable $n+1$ fois. D'après Taylor-Young, f admet un développement limité à l'ordre $n+1$, donc $f(x) = a_0 + a_1 x + \cdots + a_{n+1}x^{n+1} + o(x^{n+1})$, avec $a_k = \frac{f^{(k)}(0)}{k!}$.

Mais on sait aussi que f' est dérivable n fois, et donc admet un développement limité de la forme $f'(x) = b_0 + b_1 x + \cdots + b_n x^n + o(x^n)$, avec cette fois

$$b_k = \frac{(f')^{(k)}(0)}{k!} = \frac{f^{(k+1)}(0)}{k!} = (k+1)a_{k+1}.$$

On peut donc trouver les a_k à partir des b_k (et le contraire aussi d'ailleurs, mais en général c'est plus intéressant dans ce sens, et c'est pourquoi on parle de la méthode « d'intégration »). Il reste juste à calculer $a_0 = f(0)$ directement, et on obtient $n+1$ coefficients du développement limité de f à partir de n coefficients du développement de f'.

Voyons ça pour $f(x) = \arctan(x)$. On est bien plus à l'aise avec la dérivée $f'(x) = \frac{1}{1+x^2}$, puisque l'on a

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots + (-1)^n x^{2n} + o(x^{2n}).$$

(On déduit celui-ci du développement de $\frac{1}{1-u}$, que l'on connaît par cœur). Pour revenir à f, il suffit « d'intégrer terme à terme », c'est-à-dire que l'on a

$$f(x) = f(0) + x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^n \frac{x^{2n+1}}{2n+1} + o(x^{2n+1}).$$

Et bien sûr $f(0) = \arctan(0) = 0$.

EXEMPLE 7.15 (TANGENTE) – Il y a de nombreuses façons de calculer le développement limité de $f(x) = \tan(x)$. La méthode que nous allons présenter est particulièrement rapide. C'est un classique qui peut donner des idées dans d'autres situations.

On commence par noter que $f'(x) = 1 + \tan(x)^2 = 1 + f(x)^2$. Ainsi $f(0) = 0$ et $f'(0) = 1 + 0 = 1$. Par Taylor-Young, nous avons le développement à l'ordre 1, à savoir $f(x) = x + o(x)$.

Mais alors

$$f'(x) = 1 + f(x)^2 = 1 + (x + o(x))^2 = 1 + x^2 + o(x^2).$$

Par la méthode d'intégration, on en déduit que

$$f(x) = x + \frac{x^3}{3} + o(x^3).$$

Et on recommence :

$$\begin{aligned}f'(x) &= 1 + \left(x + \frac{x^3}{3} + o(x^3)\right)^2\\ &= 1 + x^2 + \frac{2}{3}x^4 + o(x^4).\end{aligned}$$

On intègre de nouveau :

$$f(x) = x + \frac{x^3}{3} + \frac{2}{15}x^5 + o(x^5).$$

On peut continuer comme ça pendant longtemps. Vous pouvez vérifier que l'on a

$$f(x) = x + \frac{x^3}{3} + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \frac{62}{2835}x^9 + o(x^9).$$

Il n'existe pas de formule générale pour l'ordre n.

Le minimum à savoir par cœur

voir les exercices 1237, 1239, 1240, 4019, 5426, 5427, 5429, 5430, 5432, 5433, 5434, 5437, 5438, 5439, 5440, 5441

Plus on connaît de développements limités par cœur, plus les suivants sont faciles. Dans cet esprit, voici une liste minimale de choses à savoir par cœur sous peine d'être incapable d'affronter les exercices. Les démonstrations ont toutes été données au cours de ce chapitre. Nous indiquons des moyens mnémotechniques.

- ◇ $e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!} + o(x^n)$.
- ◇ $\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + o(x^{2n+1})$.
  (On garde les termes pairs de l'exponentielle avec un signe une fois sur deux.)
- ◇ $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} + o(x^{2n})$.
  (Pareil avec les termes impairs.)
- ◇ $(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2} x^2 + \cdots + \binom{\alpha}{n} x^n + o(x^n)$
  (« Formule du binôme ».)
- ◇ $\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + o(x^n)$.
  (C'est un cas particulier de la formule précédente, mais il est tellement important qu'il faut savoir l'écrire rapidement.)
- ◇ $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots + (-1)^{n-1} \frac{x^n}{n} + o(x^n)$.
  (En dérivant on doit retrouver la formule pour $(1+x)^{-1}$.)
- ◇ Pour arctan, arcsin et arccos, on dérive et on fait un développement de la dérivée à l'aide de la formule pour $(1+x)^\alpha$.

Chapitre 8

Intégrale de Riemann

Le problème de départ que nous nous proposons de résoudre dans ce chapitre est le suivant. Étant donnée une fonction f, existe-t-il une fonction F telle que

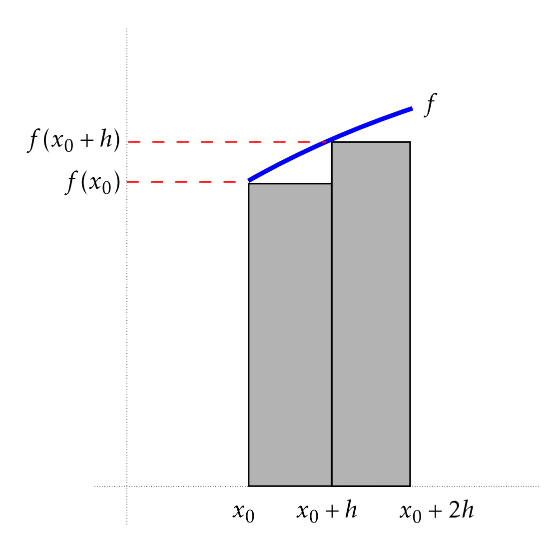
$$F' = f \quad ? \tag{*}$$

L'équation (*) est le premier exemple *d'équation différentielle* que nous rencontrons. Il s'agit d'une équation dans laquelle « l'inconnue » est une fonction, ici F, et la condition sur F fait intervenir sa dérivée. Nous verrons que résoudre (*) est une première étape importante pour résoudre des équations différentielles plus compliquées. Notons qu'une solution de (*) est appelée une *primitive* de f.

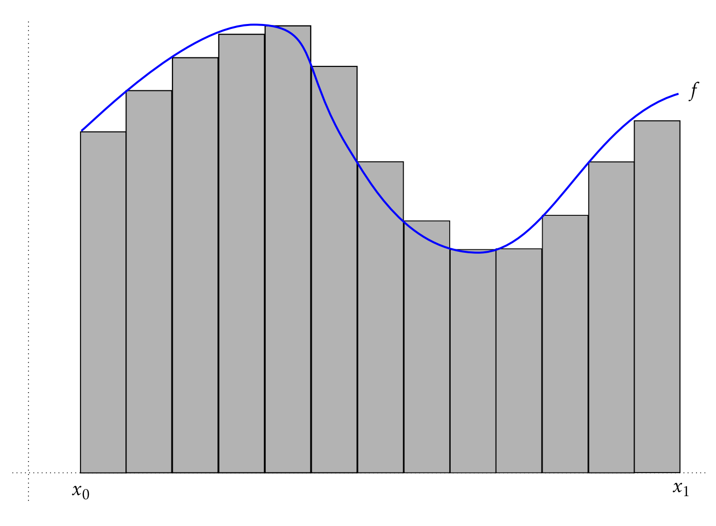
Il est facile d'étudier l'unicité des solutions, s'il y en a. En effet, supposons que f soit définie sur un *intervalle* I, et que nous cherchions F sur I. En présence de deux primitives F et G, on observe que $(F - G)' = F' - G' = f - f = 0$. La fonction $F - G$ est alors constante : deux primitives d'une même fonction sur un intervalle diffèrent d'une constante, c'est-à-dire que $G(x) = F(x) + c$ pour $x \in I$.

Le point délicat est de montrer que, sous certaines conditions, il existe au moins une primitive. On aimerait également pouvoir calculer explicitement les valeurs prises par F. Dans ce chapitre nous allons démontrer qu'une primitive existe lorsque la fonction f est continue.

La stratégie est la suivante. Supposons que $F' = f$, que $F(x_0) = 0$, et examinons la condition $F'(x_0) = f(x_0)$. Si l'on s'écarte un peu de x_0 pour atteindre le point $x_0 + h$ avec h « petit », alors $F(x_0 + h)$ est proche de $F(x_0) + F'(x_0)h = f(x_0)h$ (on remplace la fonction par son développement limité à l'ordre 1). L'idée est d'interpréter la quantité $F(x_0 + h) \simeq f(x_0)h$ comme l'aire du rectangle de hauteur $f(x_0)$ et de largeur h. De même la différence $F(x_0+2h)-F(x_0+h)$ devrait être proche de $F'(x_0+h)h = f(x_0 + h)h$, qui est l'aire du rectangle de hauteur $f(x_0 + h)$ et de largeur h. Au total $F(x_0 + 2h) = (F(x_0 + 2h) - F(x_0 + h)) + F(x_0 + h)$ est proche de la somme des aires des deux rectangles considérés sur le dessin ci-dessous.



De même pour tout entier k, la valeur $F(x_0 + kh)$ est proche de la somme des aires de k rectangles. Pour trouver une approximation de $F(x_1)$ pour $x_1 > x_0$, on peut découper l'intervalle $[x_0, x_1]$ en morceaux de largeur h, et calculer l'aire des rectangles obtenus. Lorsque h devient de plus en plus petit, intuitivement, on obtient l'aire de la zone située entre le graphe de f et l'axe des abscisses, et entre les droites verticales d'équations $x = x_0$ et $x = x_1$.



Cette analyse étant faite, pour *définir* la fonction F, il reste à définir rigoureusement ce qu'on entend par « aire ». C'est ce que nous allons faire, et nous l'appellerons *l'intégrale* de f, entre deux bornes données (sur le dessin, x_0 et x_1). Nous pourrons alors poser $F(x) =$ l'intégrale de f entre x_0 et x, et nous montrerons (assez facilement) que ce procédé donne bien une primitive de f.

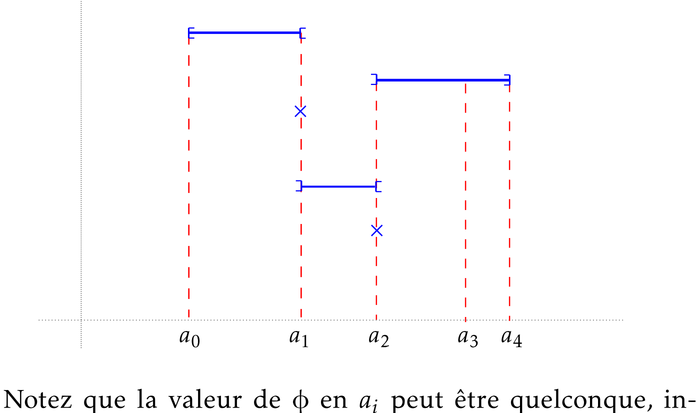
Nous obtenons au passage une application intéressante, qui est celle que vous avez étudiée en Terminale : dans de nombreux cas, c'est le calcul de l'aire qui nous intéresse, alors que l'on peut trouver une primitive directement. La relation entre « aire » et « primitive » est utile dans les deux sens.

On fixe un intervalle compact $[a,b]$. Les premières fonctions pour lesquelles on peut définir facilement ce qu'est « l'aire sous la courbe » sont les fonctions « en escaliers » :

Définition 8.1 – Une fonction ϕ sur $[a,b]$ est dite *en escaliers* lorsqu'il existe des nombres $a_0 = a < a_1 < a_2 < \cdots < a_n = b$ tels que ϕ est constante sur chaque intervalle ouvert $]a_i, a_{i+1}[$.

La famille $a = (a_0, a_1, \ldots, a_n)$ est appelée une *subdivision (adaptée à ϕ)*.

Voici un exemple de fonction en escaliers.



Notez que la valeur de ϕ en a_i peut être quelconque, indépendamment des valeurs prises sur $]a_{i-1}, a_i[$ et $]a_i, a_{i+1}[$. Sur cet exemple on voit bien que la subdivision $a = (a_0, a_1, a_2, a_3, a_4)$ n'est pas unique : en effet ϕ est en réalité constante sur $]a_2, a_4[$, et on aurait pu prendre la subdivision $a' = (a_0, a_1, a_2, a_4)$.

On peut alors poser la définition suivante :

Définition 8.2 – Soit ϕ en escaliers, et $a = (a_0, \ldots, a_n)$ une subdivision adaptée. Soit α_i la valeur de ϕ sur $]a_i, a_{i+1}[$. On pose :

$$\mathrm{I}(\phi, a) = \sum_{i=0}^{n-1} (a_{i+1} - a_i)\alpha_i \,.$$

Ce nombre est bien « l'aire » des rectangles définis par ϕ (avec cependant la possibilité que α_i soit négatif). Géométriquement on s'attend donc à ce que $\mathrm{I}(\phi)$ ne dépende pas du choix de la subdivision. C'est le cas :

Lemme 8.3 – *Le nombre $\mathrm{I}(\phi, a)$ ne dépend que de ϕ et pas du choix de la subdivision a.*

On va donc pouvoir utiliser la notation $\mathrm{I}(\phi)$.

Démonstration. Étudions un premier cas simple : si a' est obtenue à partir de a en retirant un point a_i, comme dans l'exemple ci-dessus, nous allons montrer que $\mathrm{I}(\phi, a') = \mathrm{I}(\phi, a)$. En effet, la fonction ϕ est alors constante sur $]a_{i-1}, a_{i+1}[$ de sorte que dans la somme définissant $\mathrm{I}(\phi, a)$, on a deux termes qui se simplifient :

$$(a_{i+1} - a_i)\alpha_i + (a_i - a_{i-1})\alpha_{i-1} = (a_{i+1} - a_{i-1})\alpha_{i-1} \,,$$

puisque $\alpha_{i-1} = \alpha_i$. L'égalité $\mathrm{I}(\phi, a') = \mathrm{I}(\phi, a)$ est alors claire.

En répétant l'opération, on constate que si a' est obtenue à partir de a en retirant plusieurs points, alors on a $\mathrm{I}(\phi, a') = \mathrm{I}(\phi, a)$ là encore.

Finalement, soient a et a' deux subdivisions quelconques (adaptées à ϕ). On définit une nouvelle subdivision a'' en prenant *tous* les points de a et de a', et en les rangeant dans l'ordre. Alors a est obtenue à partir de a'' en retirant les points de a', donc $\mathrm{I}(\phi, a'') = \mathrm{I}(\phi, a)$; et a' est obtenue à partir de a'' en retirant les points de a, donc $\mathrm{I}(\phi, a'') = \mathrm{I}(\phi, a')$. Finalement $\mathrm{I}(\phi, a) = \mathrm{I}(\phi, a')$. □

Lorsque ϕ et ψ sont deux fonctions (quelconques) sur $[a,b]$ telles que $\phi(x) \leq \psi(x)$ pour chaque $x \in [a,b]$, on écrira simplement $\phi \leq \psi$. La propriété suivante est très utile :

Lemme 8.4 – *Si ϕ et ψ sont en escaliers, et si $\phi \leq \psi$, alors $\mathrm{I}(\phi) \leq \mathrm{I}(\psi)$.*

Démonstration. Soit $a_0 < a_1 < \cdots < a_n$ une subdivision telle que ϕ, resp. ψ, est constante de valeur α_i, resp. α'_i, sur $]a_i, a_{i+1}[$. Grâce au lemme précédent, on peut utiliser cette subdivision (qui n'est pas forcément minimale) pour calculer les I. Par hypothèse on a $\alpha_i \leq \alpha'_i$, donc

$$\mathrm{I}(\phi) = \sum_{i=0}^{n-1} (a_{i+1} - a_i)\alpha_i \leq \sum_{i=0}^{n-1} (a_{i+1} - a_i)\alpha'_i = \mathrm{I}(\psi) \,,$$

comme annoncé. □

Il est bon de noter qu'on ne change pas $\mathrm{I}(\phi)$ si on change la fonction ϕ en un nombre fini de points. De même, si $\phi(x) \leq \psi(x)$ pour tous les x dans $[a,b]$ *sauf* pour un nombre fini de valeurs $x = x_1, \ldots, x = x_k$, alors on peut quand même conclure que $\mathrm{I}(\phi) \leq \mathrm{I}(\psi)$. Dans ce qui suit, on va utiliser ce genre de simplifications de manière implicite.

Nous arrivons à la définition de l'intégrale de Riemann. Soit f une fonction quelconque, bornée, sur $[a,b]$. On définit

$$\mathrm{I}^+(f) = \inf\{\mathrm{I}(\psi) \mid \psi \text{ en escalier telle que } \psi \geq f\}$$

et

$$\mathrm{I}^-(f) = \sup\{\mathrm{I}(\phi) \mid \phi \text{ en escalier telle que } \phi \leq f\}.$$

D'après le lemme précédent, la relation $\phi \leq f \leq \psi$ donne $\mathrm{I}(\phi) \leq \mathrm{I}(\psi)$, et par suite $\mathrm{I}^-(f) \leq \mathrm{I}^+(f)$.

Définition 8.5 – Lorsque $\mathrm{I}^-(f) = \mathrm{I}^+(f)$, on dit que f *est intégrable au sens de Riemann*, et on note

$$\int_a^b f$$

la valeur de $\mathrm{I}^{\pm}(f)$, que l'on appelle *intégrale de f sur $[a,b]$*. On note aussi parfois

$$\int_a^b f(t)\,\mathrm{d}t \,.$$

Ici la variable t est muette, et peut être remplacée par n'importe quelle lettre, souvent x ou y ou u...

Noter que la définition contient un sup, et un inf. Ceci est possible car les sup et les inf existent dans $\mathbb{R}$, comme nous l'avons vu. La théorie des intégrales atteste, à nouveau, de l'importance de cette propriété des nombres réels.

Lemme 8.6 – *Lorsque f est en escaliers, on a*

$$\int_a^b f = \mathrm{I}(f) \,.$$

Démonstration. On peut prendre $\phi = f$, et $\phi \leq f$ donne $\mathrm{I}^-(f) \geq \mathrm{I}(\phi) = \mathrm{I}(f)$. De même, en prenant $\psi = f$ on obtient $\mathrm{I}^+(f) \leq \mathrm{I}(f)$. Ainsi, on a $\mathrm{I}^-(f) \geq \mathrm{I}^+(f)$, d'où $\mathrm{I}^-(f) = \mathrm{I}^+(f) = \mathrm{I}(f)$. □

Notre premier objectif est de trouver d'autres exemples de fonctions intégrables, en dehors des fonctions en escaliers. La définition ci-dessus est très concise, et montre de manière explicite l'utilisation de sup et de inf, mais pour montrer concrètement qu'une fonction donnée est intégrable on va utiliser le critère simple suivant.

LEMME 8.7 – *Soit f une fonction sur $[a,b]$. Alors f est intégrable $\iff$ il existe deux suites $(\phi_n)_{n\geq 0}$ et $(\psi_n)_{n\geq 0}$ de fonctions en escaliers telles que $\phi_n \leq f \leq \psi_n$, et telles que*

$$\lim_{n\to+\infty} \int_a^b \psi_n - \int_a^b \phi_n = 0 .$$

On a alors

$$\int_a^b f = \lim_{n\to\infty} \int_a^b \phi_n = \lim_{n\to\infty} \int_a^b \psi_n .$$

Démonstration. Commençons par $\Longleftarrow$. De $\phi_n \leq f$ on tire $I^-(f) \geq \int_a^b \phi_n$ par définition ; de même $I^+(f) \leq \int_a^b \psi_n$. On tire

$$0 \leq I^+(f) - I^-(f) \leq \int_a^b \psi_n - \int_a^b \phi_n .$$

En passant à la limite sur n, on obtient bien $I^+(f) = I^-(f)$, c'est-à-dire que f est intégrable.

Maintenant regardons $\Longrightarrow$. Soit n un entier ≥ 1. Comme $I^-(f) - \frac{1}{n} < I^-(f)$, on sait par définition du sup (= le plus petit des majorants d'un ensemble) qu'il existe une fonction en escaliers ϕ_n telle que $\phi_n \leq f$ et telle que $I(\phi_n) \geq I^-(f) - \frac{1}{n}$. De la même manière, il existe ψ_n en escaliers telle que $\psi_n \geq f$ et telle que $I(\psi_n) \leq I^+(f) + \frac{1}{n}$. Si f est intégrable on a $I^+(f) = I^-(f)$ et donc

$$0 \leq \int_a^b \psi_n - \int_a^b \phi_n \leq \left(I^+(f) + \frac{1}{n}\right) - \left(I^-(f) - \frac{1}{n}\right) = \frac{2}{n}$$

(la première inégalité provient du lemme 8.4). En passant à la limite, on obtient le résultat.

Pour finir, de l'inégalité $\phi_n \leq f \leq \psi_n$, on tire

$$\int_a^b \phi_n \leq \int_a^b f \leq \int_a^b \psi_n$$

par définition même de l'intégrale. En écrivant ceci sous la forme

$$0 \leq \int_a^b f - \int_a^b \phi_n \leq \int_a^b \psi_n - \int_a^b \phi_n$$

et en faisant tendre n vers l'infini, on obtient bien la formule souhaitée. □

Par exemple on peut utiliser ce résultat pour montrer :

PROPOSITION 8.8 – *Soit f une fonction monotone sur $[a,b]$. Alors f est intégrable.*

Démonstration. On va faire la preuve dans le cas où f est croissante. Le cas où elle est décroissante est similaire.

Soit n un entier ≥ 1. On va découper $[a,b]$ en n morceaux en posant $a_i = a + i\left(\frac{b-a}{n}\right)$, pour $0 \leq i \leq n$. On définit maintenant deux fonctions en escaliers ϕ_n, resp. ψ_n, qui sont constantes sur chaque intervalle $]a_i, a_{i+1}[$ de valeur $f(a_i)$, resp. $f(a_{i+1})$. Par croissance de f, on a $\phi_n \leq f \leq \psi_n$. Calculons maintenant $\int_a^b \phi_n = I(\phi_n)$ directement avec la formule définissant I : comme $a_{i+1} - a_i = \frac{b-a}{n}$, on obtient

$$\int_a^b \phi_n = \frac{b-a}{n} \sum_{i=0}^{n-1} f(a_i).$$

De même

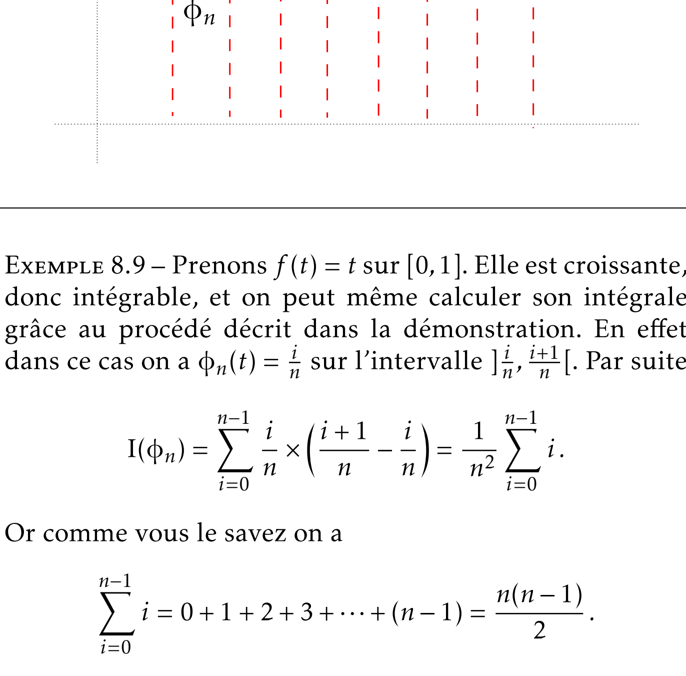
$$\int_a^b \psi_n = \frac{b-a}{n} \sum_{i=0}^{n-1} f(a_{i+1}).$$

On tire

$$\int_a^b \psi_n - \int_a^b \phi_n = \frac{f(b) - f(a)}{n}(b-a).$$

On fait ensuite tendre n vers l'infini, et les conditions du lemme 8.7 sont réunies. Donc f est intégrable. □

Cette démonstration est illustrée sur le dessin suivant.



EXEMPLE 8.9 – Prenons $f(t) = t$ sur $[0,1]$. Elle est croissante, donc intégrable, et on peut même calculer son intégrale grâce au procédé décrit dans la démonstration. En effet dans ce cas on a $\phi_n(t) = \frac{i}{n}$ sur l'intervalle $]\frac{i}{n}, \frac{i+1}{n}[$. Par suite

$$I(\phi_n) = \sum_{i=0}^{n-1} \frac{i}{n} \times \left(\frac{i+1}{n} - \frac{i}{n}\right) = \frac{1}{n^2} \sum_{i=0}^{n-1} i .$$

Or comme vous le savez on a

$$\sum_{i=0}^{n-1} i = 0 + 1 + 2 + 3 + \cdots + (n-1) = \frac{n(n-1)}{2} .$$

Ceci donne

$$I(\phi_n) = \frac{1}{2} - \frac{1}{2n} ,$$

et

$$\int_0^1 t \, dt = \lim_{n\to\infty} I(\phi_n) = \frac{1}{2} .$$

EXEMPLE 8.10 – Prenons maintenant $f(t) = e^t$ sur $[0,1]$. Là encore f est croissante donc intégrable, et cette fois on a (toujours avec les notations de la démonstration) $\phi_n(t) = e^{\frac{i}{n}}$ sur $]\frac{i}{n}, \frac{i+1}{n}[$. Ici

$$I(\phi_n) = \sum_{i=0}^{n-1} e^{\frac{i}{n}} \times \left(\frac{i+1}{n} - \frac{i}{n}\right) = \frac{1}{n} \sum_{i=0}^{n-1} e^{\frac{i}{n}} .$$

Si on pose $\alpha = e^{\frac{1}{n}}$, on a

$$\sum_{i=0}^{n-1} \left(e^{\frac{1}{n}}\right)^i = 1 + \alpha + \alpha^2 + \cdots + \alpha^{n-1} = \frac{1-\alpha^n}{1-\alpha} = \frac{1-e}{1-e^{\frac{1}{n}}} .$$

Utilisons le développement limité $e^u = 1 + u + o(u)$. On en tire $e^{\frac{1}{n}} = 1 + \frac{1}{n} + o(\frac{1}{n})$, d'où

$$I(\phi_n) = \frac{1}{n} \cdot \frac{1-e}{-\frac{1}{n} + o(\frac{1}{n})} = \frac{1-e}{-1 + o(1)} .$$

En passant à la limite, on obtient

$$\int_0^1 e^t \, dt = e - 1 .$$

On a donc déjà un grand nombre d'exemples de fonctions qui sont intégrables, comme l'exponentielle, le logarithme, etc. Noter que l'on ne suppose même pas que f est continue dans la proposition.

Pour ce qui est des autres fonctions « usuelles », comme le sinus et le cosinus par exemple, on peut remarquer que sur un intervalle compact $[a,b]$, même si elles ne sont pas forcément monotones, on peut découper $[a,b]$ en un nombre *fini* de sous-intervalles sur lesquels ces fonctions sont monotones. Les propriétés élémentaires des intégrales, que nous allons voir tout de suite, vont alors montrer que ces fonctions sont également intégrables.

Commençons par la « relation de Chasles » :

Proposition 8.11 – *Soit f une fonction sur $[a,b]$, et soit $x \in [a,b]$. Alors f est intégrable sur $[a,b] \iff f$ est intégrable sur $[a,x]$ et sur $[x,b]$.*

De plus, on a la relation dite de Chasles :

$$\int_a^b f = \int_a^x f + \int_x^b f.$$

Le schéma de la démonstration est à retenir : on va le réutiliser constamment.

Démonstration. On commence par traiter le cas où f est en escaliers. Alors f est intégrable partout, donc il y a seulement la relation de Chasles à montrer. En utilisant la formule pour I, on s'aperçoit que le résultat est évident.

On passe au cas général. Montrons $\implies$. On suppose que f est intégrable sur $[a,b]$, et on prend ϕ_n et ψ_n comme dans le lemme 8.7. La relation $\phi_n \le f \le \psi_n$ étant valable sur $[a,b]$, elle est aussi valable sur $[a,x]$ et $[x,b]$. De plus, comme ϕ_n et ψ_n sont en escaliers et que l'on a une relation de Chasles dans ce cas, on a :

$$\int_a^b \psi_n - \int_a^b \phi_n = \left(\int_a^x \psi_n - \int_a^x \phi_n \right) + \left(\int_x^b \psi_n - \int_x^b \phi_n \right).$$

Les termes entre parenthèses sont ≥ 0 car $\phi_n \le \psi_n$ et on a le lemme 8.4.

En particulier on a

$$0 \le \int_a^x \psi_n - \int_a^x \phi_n \le \int_a^b \psi_n - \int_a^b \phi_n$$

et en faisant tendre n vers $+\infty$, on obtient

$$\lim_{n \to +\infty} \int_a^x \psi_n - \int_a^x \phi_n = 0.$$

Donc le lemme 8.7 (dans le sens $\Longleftarrow$!) nous dit que f est intégrable sur $[a,x]$. De même entre x et b.

Montrons maintenant $\Longleftarrow$. Notons f_1 la restriction de f à $[a,x]$. On suppose f_1 intégrable, donc on peut choisir des suites $\phi_{n,1}$ et $\psi_{n,1}$ comme dans le lemme 8.7. De même, on définit f_2, $\phi_{n,2}$ et $\psi_{n,2}$ sur $[x,b]$. On recolle les morceaux, en définissant ϕ_n sur $[a,b]$ par $\phi_n(t) = \phi_{n,1}(t)$ si $t \in [a,x]$, et $\phi_n(t) = \phi_{n,2}(t)$ si $t \in]x,b]$; de même on définit ψ_n.

On veut utiliser encore le lemme 8.7 pour conclure. Vérifions les hypothèses : ϕ_n et ψ_n sont en escaliers ; on a bien $\phi_n \le f \le \psi_n$, car on peut vérifier ceci séparément sur $[a,x]$ et sur $[x,b]$ et on a supposé que $\phi_{n,i} \le f \le \psi_{n,i}$ pour $i = 1$ ou 2 ; enfin, la relation de Chasles pour les fonctions en escaliers donne encore :

$$\int_a^b \psi_n - \int_a^b \phi_n = \left(\int_a^x \psi_n - \int_a^x \phi_n \right) + \left(\int_x^b \psi_n - \int_x^b \phi_n \right).$$

Le terme dans la première parenthèse est en fait

$$\int_a^x \psi_{n,1} - \int_a^x \phi_{n,1}$$

puisque par construction, $\phi_n = \phi_{n,1}$ sur $[a,x]$; de même le terme dans la deuxième parenthèse peut s'écrire avec $\psi_{n,2}$ et $\phi_{n,2}$. Donc les termes entre parenthèses tendent vers 0 quand n tend vers $+\infty$, en on a finalement réuni toutes les hypothèses du lemme 8.7, qui nous dit que f est intégrable sur $[a,b]$.

De plus, on sait aussi que

$$\int_a^b f = \lim_{n \to +\infty} \int_a^b \phi_n, \text{ que } \int_a^x f = \int_a^x f_1 = \lim_{n \to +\infty} \int_a^x \phi_{n,1}$$

et que

$$\int_x^b f = \int_x^b f_2 = \lim_{n \to +\infty} \int_x^b \phi_{n,2}.$$

En écrivant la relation de Chasles pour ϕ_n (qui est en escaliers) et en passant à la limite, on obtient la relation de Chasles pour f quelconque. □

Exemple 8.12 – Supposons que f soit une fonction sur $[a,b]$, et qu'il existe une subdivision $a_0 = a < a_1 < \cdots < a_n = b$ telle que f est monotone sur $[a_i : a_{i+1}]$ (par exemple les fonctions sinus et cosinus ont cette propriété sur n'importe quel intervalle compact). Alors d'après la proposition 8.8, f est intégrable sur $[a_i : a_{i+1}]$. En utilisant la proposition 8.11 (plusieurs fois !), on obtient que f est intégrable sur $[a,b]$.

On a donc montré que toutes les fonctions dites « usuelles » sont intégrables. Par contre nous n'avons encore rien dit sur la façon de calculer les intégrales, bien sûr.

Toutes les autres propriétés élémentaires de l'intégrale se démontrent de la même manière. On a principalement :

Proposition 8.13 – *Propriétés de l'intégrale :*

1. *(Linéarité) Si f et g sont intégrables, et si $\lambda, \mu \in \mathbb{R}$, alors $\lambda f + \mu g$ est intégrable et on a :*

$$\int_a^b (\lambda f(t) + \mu g(t))\,dt = \lambda \int_a^b f(t)\,dt + \mu \int_a^b g(t)\,dt.$$

En d'autres termes, l'ensemble E *des fonctions intégrables est un espace vectoriel, et l'intégrale est une application linéaire* $\mathrm{E} \to \mathbb{R}$.

2. *(Croissance) Si f et g sont intégrables, et si $f \le g$, alors*

$$\int_a^b f \le \int_a^b g.$$

Esquisse. Le principe est le même que pour la proposition précédente : on vérifie les propriétés pour les fonctions en escaliers d'abord (par exemple le point 2 est donné dans ce cas par le lemme 8.4), et on utilise le lemme 8.7 pour montrer par passage à la limite que les propriétés sont en fait vérifiées pour des fonctions quelconques. □

Sur le même modèle, nous allons conclure en établissant la très utile « inégalité triangulaire » pour les intégrales. Nous aurons besoin des définitions suivantes : si f est une fonction quelconque, on pose

$$f^+(x) = \begin{cases} f(x) & \text{si } f(x) \ge 0 \\ 0 & \text{sinon,} \end{cases}$$

$$f^-(x) = \begin{cases} -f(x) & \text{si } f(x) \le 0 \\ 0 & \text{sinon.} \end{cases}$$

On vérifie que l'on a $f(x) = f^+(x) - f^-(x)$ et aussi $|f(x)| = f^+(x) + f^-(x)$. Aussi, on note que si $f \le g$, alors $f^+ \le g^+$ et $g^- \le f^-$.

Proposition 8.14 – *Si f est intégrable sur $[a,b]$, alors $|f|$ est également intégrable et on a l'inégalité dite triangulaire*

$$\left| \int_a^b f \right| \le \int_a^b |f|.$$

Démonstration. On commence par montrer que f^+ est intégrable. Puisque f est intégrable, on peut trouver ϕ_n et ψ_n comme dans le lemme 8.7. Les fonctions ϕ_n^+ et ψ_n^+ sont en escaliers, et on a $\phi_n^+ \le f^+ \le \psi_n^+$. On vérifie (exercice) que $\psi_n^+ - \phi_n^+ \le \psi_n - \phi_n$. Donc finalement

$$0 \le \int_a^b \psi_n^+ - \int_a^b \phi_n^+ \le \int_a^b \psi_n - \int_a^b \phi_n.$$

(On a utilisé la linéarité de l'intégrale.) En passant à la limite, on voit que l'on peut appliquer le lemme 8.7 dans le sens $\Longleftarrow$, et on conclut que f^+ est intégrable.

De même, on montre que f^- est intégrable. Donc $|f| = f^+ + f^-$ est intégrable, comme somme de fonctions intégrables.

Enfin, de $f \le |f|$ on tire

$$\int_a^b f \le \int_a^b |f|$$

par croissance de l'intégrale. De même de $-f \le |f|$ on tire

$$-\int_a^b f \le \int_a^b |f|.$$

On a bien prouvé l'inégalité triangulaire. □

Nous allons maintenant montrer le résultat théorique principal du chapitre, à savoir l'intégrabilité des fonctions continues.

THÉORÈME 8.15 – *Soit f une fonction continue sur $[a,b]$. Alors f est intégrable.*

Démonstration. On commence comme dans la proposition 8.8. Soit n un entier ≥ 1. On va découper $[a,b]$ en n morceaux en posant $a_i = a + i\left(\frac{b-a}{n}\right)$, pour $0 \leq i \leq n$. On définit maintenant deux fonctions en escaliers ϕ_n, resp. ψ_n, qui sont constantes sur chaque intervalle $]a_i, a_{i+1}[$ de valeur α_i, resp. β_i, avec

$$\alpha_i = \min\{f(t) \mid t \in [a_i, a_{i+1}]\}$$

et

$$\beta_i = \max\{f(t) \mid t \in [a_i, a_{i+1}]\}.$$

(Ceci a un sens d'après le corollaire 5.4, qui garantit que $\beta_i \neq +\infty$ par exemple.) Ainsi $\phi_n \leq f \leq \psi_n$.

Soit $\varepsilon > 0$. Comme f est uniformément continue (d'après le théorème de Heine 5.11), il existe $\delta > 0$ tel que, si x et y sont dans $[a,b]$ et vérifient $|x-y| < \delta$, alors $|f(x) - f(y)| < \varepsilon$. Mais pour n suffisamment grand, on a $\frac{b-a}{n} < \delta$; fixons un tel n. Comme $a_{i+1} - a_i = \frac{b-a}{n}$, on voit que si x et y sont pris tous les deux dans $]a_i, a_{i+1}[$, alors $|f(x) - f(y)| \leq \varepsilon$. On en conclut que $\beta_i - \alpha_i \leq \varepsilon$ pour chaque i.

Si on calcule maintenant $\int_a^b (\psi_n - \phi_n)$ directement avec la formule pour I, on obtient :

$$\int_a^b (\psi_n - \phi_n) = \sum_{i=0}^{n-1} (a_{i+1} - a_i)(\beta_i - \alpha_i) \leq (b-a)\frac{\varepsilon}{n} + \frac{\varepsilon}{n} + \cdots + \frac{\varepsilon}{n} = \varepsilon(b-a).$$

Ainsi, pour tout $\varepsilon > 0$, on a $0 \leq \int_a^b (\psi_n - \phi_n) \leq \varepsilon(b-a)$ dès que n est suffisamment grand ; c'est dire que

$$\lim_{n\to+\infty} \int_a^b (\psi_n - \phi_n) = 0.$$

Le lemme 8.7 nous dit alors que f est intégrable. □

On s'aperçoit *a posteriori* qu'on aurait pu considérer d'autres fonctions en escaliers dans cette démonstration, et obtenir la même limite. Pour être précis, on a le résultat suivant :

PROPOSITION 8.16 – *Soit f une fonction continue sur $[a,b]$. Pour chaque entier $n \geq 1$, on pose $a_i = a + i\left(\frac{b-a}{n}\right)$ pour $0 \leq i \leq n$, et on choisit $x_{i,n} \in [a_i, a_{i+1}]$. Alors :*

$$\lim_{n\to+\infty} \frac{b-a}{n} \sum_{i=0}^{n-1} f(x_{i,n}) = \int_a^b f(t)\,dt.$$

Démonstration. On reprend les notations de la démonstration du théorème 8.15.

Soit alors θ_n la fonction en escaliers, constante sur $]a_i, a_{i+1}[$, de valeur $f(x_{i,n})$. La formule pour I montre que $I(\theta_n)$ est bien la somme dont on cherche la limite.

Par définition, on a alors $\phi_n \leq \theta_n \leq \psi_n$, et donc

$$I(\phi_n) \leq I(\theta_n) \leq I(\psi_n).$$

Les termes de gauche et de droite de cette inégalité tendent vers $\int_a^b f$ comme on l'a observé, et donc le terme du milieu converge vers la même limite. □

Les sommes de cette forme sont appelées « sommes de Riemann ». Comme on vient de le voir dans la démonstration, elles expriment l'intégrale de certaines fonctions en escaliers qui s'approchent de f, et la proposition affirme qu'à la limite on obtient l'intégrale de f. Il est clair qu'on aurait pu trouver des fonctions en escaliers plus générales pour lesquelles le résultat est encore vrai (notamment avec des subdivisions de l'intervalle $[a,b]$ plus compliquées) et c'est parfois sous cette forme que l'on énonce un résultat sur les sommes de Riemann dans certains livres. D'une manière générale, lorsqu'on étudie certaines sommes, il faut garder en tête l'idée d'utiliser des fonctions en escaliers et des intégrales.

En pratique ceci dit, même la version ci-dessus est trop générale. Dans presque tous les cas que l'on rencontre, le choix est $x_{i,n} = a_i$; et très souvent, on a même $a = 0$ et $b = 1$. Ainsi, il est bon de mémoriser la formule suivante : lorsque f est continue sur $[0,1]$, on a

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right) = \int_0^1 f(t)\,dt.$$

EXEMPLE 8.17 – Soit

$$S_n = \sum_{i=0}^{n-1} \frac{n}{i^2 + n^2}.$$

La suite $(S_n)_{n\geq 1}$ a-t-elle une limite ? Si on pense à utiliser une somme de Riemann, il faut trouver une bonne fonction f. Commençons par faire apparaître le $\frac{1}{n}$, puis mettons la quantité $\frac{i}{n}$ en évidence :

$$S_n = \frac{1}{n} \sum_{i=0}^{n-1} \frac{n^2}{i^2 + n^2} = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{1 + \frac{i^2}{n^2}}.$$

On a donc bien envie d'introduire la fonction f définie sur $[0,1]$ par $f(t) = \frac{1}{1+t^2}$, puisqu'alors :

$$S_n = \frac{1}{n} \sum_{i=0}^{n-1} f\left(\frac{i}{n}\right).$$

Donc S_n est une somme de Riemann, et

$$\lim_{n\to\infty} S_n = \int_0^1 f(t)\,dt = \int_0^1 \frac{dt}{1+t^2}.$$

Reste à calculer cette intégrale ! Nous allons maintenant voir comment.

voir les exercices (sans calculer les intégrales) 2100 et 5446 questions 1, 6, 7 puis 5, 8

Il existe une hiérarchie des fonctions sur $[a,b]$:

$$f \text{ dérivable} \Longrightarrow f \text{ continue} \Longrightarrow f \text{ intégrable.}$$

On a déjà vu la première implication ; la deuxième est le contenu du théorème 8.15. On va maintenant montrer qu'en intégrant une fonction, on remonte dans la hiérarchie.

Soit donc f intégrable sur $[a,b]$. On note

$$F(x) = \int_a^x f(t)\,dt.$$

D'après la proposition 8.11, F est bien définie.

Proposition 8.18 – *Pour toute fonction f intégrable, la fonction F est continue.*

Démonstration. Soit donc $x_0 \in [a,b]$. D'après la relation de Chasles, on a :

$$F(x) - F(x_0) = \int_{x_0}^x f(t)\,dt.$$

La fonction f est bornée par hypothèse : écrivons $|f(t)| \le M$ pour $t \in [a,b]$. Utilisons maintenant l'inégalité triangulaire :

$$|F(x) - F(x_0)| \le \int_{x_0}^x |f(t)|\,dt \le \int_{x_0}^x M\,dt = M(x - x_0).$$

(On a supposé $x \ge x_0$, le cas $x \le x_0$ est similaire.) Ceci montre que

$$\lim_{x \to x_0} F(x) - F(x_0) = 0,$$

c'est-à-dire que F est continue en x_0. □

Dans le même genre, on a le résultat annoncé depuis l'introduction de ce chapitre :

Proposition 8.19 – *Si f est continue, F est dérivable. De plus, on a $F' = f$.*

Démonstration. Soit $x_0 \in [a,b]$. En utilisant la relation

$$\int_{x_0}^x f(x_0)\,dt = f(x_0)(x - x_0)$$

et la relation de Chasles, on peut écrire :

$$\frac{F(x) - F(x_0)}{x - x_0} - f(x_0) = \frac{1}{x - x_0}\int_{x_0}^x (f(t) - f(x_0))\,dt.$$

Soit $\varepsilon > 0$. Comme f est continue en x_0, il existe $\delta > 0$ tel que $|x - x_0| < \delta$ entraîne $|f(x) - f(x_0)| < \varepsilon$. Fixons un tel x, et pour simplifier disons $x > x_0$. Si maintenant $t \in [x_0, x]$, on a aussi $|t - x_0| < \delta$ et donc $|f(t) - f(x_0)| < \varepsilon$. On en déduit :

$$\begin{aligned}
\left| \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) \right| &\le \frac{1}{x - x_0}\int_{x_0}^x |f(t) - f(x_0)|\,dt \\
&\le \frac{1}{x - x_0}\int_{x_0}^x \varepsilon\,dt \\
&\le \frac{\varepsilon(x - x_0)}{x - x_0} = \varepsilon.
\end{aligned}$$

On a donc bien

$$\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} - f(x_0) = 0.$$

Donc F est dérivable en x_0, et $F'(x_0) = f(x_0)$. □

On a donc bien démontré, comme annoncé dans l'introduction, que toute fonction continue f admet une primitive, ici notée F. Notons que le point délicat était vraiment de montrer que f est intégrable : le reste de la preuve ci-dessus n'utilise que des propriétés élémentaires de l'intégrale.

La proposition a une conséquence que l'on appelle, en toute simplicité, le « théorème fondamental de l'analyse » :

Théorème 8.20 – *Si f est une fonction continument dérivable sur $[a,b]$, alors*

$$\int_a^b f'(t)\,dt = f(b) - f(a).$$

Démonstration. C'est une reformulation de la proposition précédente. Dire que f est continument dérivable signifie que f' est continue, on peut donc appliquer la proposition 8.19 avec f' a la place de f. On en déduit que la fonction définie par

$$g(x) = \int_a^x f'(t)\,dt$$

est dérivable de dérivée f'. Donc g a la même dérivée que f, et ceci entraîne $f(x) = g(x) + c$ pour une constante c. En évaluant en a, on obtient $f(a) = g(a) + c = c$, d'où $f(x) = g(x) + f(a)$. En $x = b$, on obtient $f(b) - f(a) = g(b)$, ce qu'on voulait. □

Exemple 8.21 – Revenons à l'intégrale rencontrée à la fin de l'exemple 8.17 :

$$\int_0^1 \frac{dt}{1+t^2} = ?$$

Pour utiliser le théorème 8.20, il nous faut trouver une fonction f telle que

$$f'(t) = \frac{1}{1+t^2}.$$

Passant en revue les dérivées des fonctions que l'on connait, on constate que l'on peut prendre

$$f(t) = \arctan(t)$$

(et les autres primitives sur $[0,1]$ sont donc de la forme $t \mapsto \arctan(t) + c$). Donc

$$\int_0^1 \frac{dt}{1+t^2} = \arctan(1) - \arctan(0) = \frac{\pi}{4}.$$

En particulier, grâce à l'exemple 8.17, on a une formule pour calculer π :

$$\pi = \lim_{n\to\infty} 4 \sum_{i=0}^{n-1} \frac{n}{i^2 + n^2}.$$

La convergence est très lente. Il faut attendre $n = 119$ pour avoir 3 chiffres corrects $(3,149984\ldots)$ et pour $n = 100\,000$ on obtient $3,141602\ldots$, donc seulement 4 chiffres corrects.

Exemple 8.22 – On peut retrouver les résultats obtenus « à la main » dans les exemples 8.9 et 8.10 très facilement. En effet la fonction $t \mapsto t$ admet pour primitive $t \mapsto \frac{t^2}{2}$, d'où

$$\int_0^1 t\,dt = \left[\frac{t^2}{2}\right]_0^1 = \frac{1}{2}.$$

Ici on a utilisé la notation avec les crochets qui vous est familière depuis la terminale, à savoir

$$[f(t)]_a^b = f(b) - f(a).$$

De même on a

$$\int_0^1 e^t\,dt = \left[e^t\right]_0^1 = e - 1.$$

On retrouve bien les mêmes valeurs.

reprendre les exercices précédents et calculer les intégrales

Exemple 8.23 – La méthode des primitives pour calculer les sommes est très puissante ; c'est précisément pour cela que les intégrales sont plus faciles à calculer que les sommes. Voyons une application. On considère la suite $(H_n)_{n\ge1}$ définie par

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{i=1}^n \frac{1}{i}.$$

(On appelle parfois (H_n) la « série harmonique », d'où la notation.) Nous allons montrer que H_n tend vers $+\infty$. À partir de la définition, c'est difficile.

Essayons donc d'utiliser des intégrales. Nous n'avons pas une somme de Riemann, mais nous pouvons malgré tout essayer d'introduire certaines fonctions en escaliers. En l'occurrence, considérons l'intervalle $[1,n]$ et la fonction en escaliers ϕ_n qui vaut $\frac{1}{i}$ sur l'intervalle $]i, i+1[$. En comptant l'aire des rectangles, on constate que

$$\int_1^n \phi_n = H_n.$$

Soit maintenant $f(t) = \frac{1}{t}$ sur le même intervalle. On a alors $\phi_n \ge f$, ce qui est une conséquence du fait que f est décroissante. Par suite

$$\int_1^n \phi_n \ge \int_1^n f.$$

Mais grâce à la méthode des primitives, nous pouvons facilement calculer l'intégrale de f : en effet le logarithme ln vérifie $\ln' = f$. Ainsi

$$\int_1^n \frac{dt}{t} = \ln(n) - \ln(1) = \ln(n).$$

Or nous savons que $\ln(n) \to +\infty$ lorsque $n \to +\infty$, donc l'inégalité $H_n \ge \ln(n)$ que nous venons d'obtenir garantit que $H_n \to +\infty$.

Une autre conséquence plus ou moins immédiate de la proposition 8.19 est la « formule du changement de variables » qui est très pratique dans les calculs :

Proposition 8.24 – *Soit u une fonction continument dérivable sur $[a,b]$, et supposons que l'image de $[a,b]$ par u soit l'intervalle $[u(a),u(b)]$. Soit f une fonction continue sur $[u(a),u(b)]$. Alors :*

$$\int_a^b f(u(t))\,u'(t)\,\mathrm{d}t = \int_{u(a)}^{u(b)} f(x)\,\mathrm{d}x\,.$$

Démonstration. Pour $\mathrm{X} \in [u(a),u(b)]$, soit

$$\mathrm{F}(\mathrm{X}) = \int_{u(a)}^{\mathrm{X}} f(x)\,\mathrm{d}x$$

et soit $g(\mathrm{T}) = \mathrm{F}(u(\mathrm{T}))$ pour $\mathrm{T} \in [a,b]$. Puisque f est continue, la proposition 8.19 permet de calculer la dérivée de F, à savoir $\mathrm{F}' = f$; et la formule pour la dérivée d'une fonction composée donne au total :

$$g'(\mathrm{T}) = \mathrm{F}'(u(\mathrm{T}))\,u'(\mathrm{T}) = f(u(\mathrm{T}))\,u'(\mathrm{T})\,.$$

D'un autre côté, soit

$$h(\mathrm{T}) = \int_a^{\mathrm{T}} f(u(t))u'(t)\,\mathrm{d}t$$

pour $\mathrm{T} \in [a,b]$. L'expression $t \to f(u(t))u'(t)$ est bien continue, donc la proposition 8.19 (encore) donne $h'(\mathrm{T}) = f(u(\mathrm{T}))u'(\mathrm{T})$.

Ainsi, les fonctions g et h ont la même dérivée. Elles sont toutes les deux nulles en $\mathrm{T} = a$, donc elles sont finalement égales. En écrivant $g(b) = h(b)$, on obtient la formule. □

Pour retenir correctement la formule, nous suggérons un petit abus de notation. Dans le membre de droite de la formule du changement de variables, l'écriture $f(x)\mathrm{d}x$ peut être évidemment remplacée par $f(y)\mathrm{d}y$ ou $f(z)\mathrm{d}z$. Si vous pensez que l'on peut utiliser *n'importe quelle lettre*, demandez-vous quand même ce que donnerait la formule avec la lettre f ou pire, avec d. Nous proposons d'utiliser la lettre u, et donc de mémoriser

$$\int_a^b f(u(t))\,u'(t)\,\mathrm{d}t = \int_{u(a)}^{u(b)} f(u)\,\mathrm{d}u\,.$$

À strictement parler, c'est un abus de notation, puisque la lettre u désigne déjà une certaine fonction. Mais l'intuition derrière ce choix est assez correcte. Si on lit la formule « de la droite vers la gauche », on constate que « en faisant varier u », donc en remplaçant u par $u(t)$, on doit remplacer $\mathrm{d}u$ par $u'(t)\mathrm{d}t$, ce qui est cohérent avec la notation $u'(t) = \mathrm{d}u/\mathrm{d}t$ utilisée en Physique.

La plupart du temps, on utilise cependant la formule de la « gauche vers la droite ». Dans ce sens, on retient que l'expression $u(t)$, éventuellement très compliquée, peut être remplacée simplement par u pour peu que l'on ait quelque part un $u'(t)\mathrm{d}t$, qui va devenir $\mathrm{d}u$. Voici un exemple.

Exemple 8.25 – Calculons

$$\int_0^1 \frac{\mathrm{d}t}{1+e^t}\,.$$

On va poser $u(t) = e^t$, pour voir. Il nous faudrait un $u'(t)\mathrm{d}t$ quelque part ; or $u'(t) = e^t$. On peut donc artificiellement écrire

$$\int_0^1 \frac{\mathrm{d}t}{1+e^t} = \int_0^1 \frac{e^t\,\mathrm{d}t}{e^t(1+e^t)} = \int_0^1 \frac{u'(t)\,\mathrm{d}t}{u(t)(1+u(t))} = \int_1^e \frac{\mathrm{d}u}{u(1+u)}\,.$$

On va pouvoir finir ce calcul facilement. Avant de le faire, remarquons que si les choses se sont bien passées, c'est surtout parce que $u'(t)$ *peut s'exprimer facilement en termes de* $u(t)$. C'est la grande qualité que l'on cherche dans un changement de variables. Faites l'expérience suivante : essayez le changement de variables $v(t) = 1/(1+e^t)$. Vous vous rendrez compte que pour faire apparaître le $v'(t)\mathrm{d}t$, on est amené à exprimer $v'(t)$ en fonction de $v(t)$. C'est faisable, mais moins facile que pour u ; au total on finit avec la même expression, mais après bien plus d'efforts.

Terminons tout de même. On écrit

$$\frac{1}{u(1+u)} = \frac{1}{u} - \frac{1}{1+u}\,.$$

Vous vérifierez cette égalité sans peine ; au chapitre suivant nous verrons comment systématiser ce genre d'astuce. Maintenant il vient :

$$\begin{aligned}\int_1^e \frac{\mathrm{d}u}{u(1+u)} &= \int_1^e \frac{\mathrm{d}u}{u} - \int_1^e \frac{\mathrm{d}u}{1+u}\\ &= [\ln(u)]_1^e - [\ln(1+u)]_1^e\\ &= 1 + \ln(2) - \ln(1+e)\,.\end{aligned}$$

voir les exercices 812, 2098, 6865, 2099 (le chapitre suivant peut aider pour certaines primitives)

– Fin de la première lecture –

Définition 8.26 – Soit $f: [a,b] \to \mathbb{R}^r$ une fonction, et écrivons

$$f(t) = (f_1(t), \dots, f_r(t)).$$

On dit que f est *Riemann-intégrable* lorsque chaque fonction f_k est Riemann-intégrable (au sens déjà défini dans ce chapitre). De plus, *l'intégrale* de f est le vecteur

$$\int_a^b f = \left(\int_a^b f_1, \int_a^b f_2, \dots, \int_a^b f_r \right).$$

Par exemple, si nous identifions comme d'habitude l'ensemble $\mathbb{C}$ des nombres complexes avec $\mathbb{R}^2$, en voyant $x+iy$ comme (x,y), alors cette définition signifie que nous avons la convention suivante : pour une fonction $f: [a,b] \to \mathbb{C}$ de la forme $f(t) = x(t) + iy(t)$, son intégrale est

$$\int_a^b f(t)\,dt = \int_a^b x(t)\,dt + i \int_a^b y(t)\,dt.$$

C'est évidemment la définition la plus naturelle.

Proposition 8.27 – *On a les propriétés suivantes.*

1. *Si f et g sont intégrables, à valeurs dans $\mathbb{R}^r$, et si λ et μ sont des constantes réelles, alors $\lambda f + \mu g$ est intégrable et*

$$\int_a^b (\lambda f(t) + \mu g(t))\,dt = \lambda \int_a^b f(t)\,dt + \mu \int_a^b g(t)\,dt.$$

De plus, si $v \in \mathbb{R}^r$ est un vecteur (constant), et si $f: [a,b] \to \mathbb{R}$ est intégrable, alors la fonction $t \mapsto f(t)v$ est intégrable et

$$\int_a^b f(t)\,v\,dt = \left(\int_a^b f(t)\,dt \right) v.$$

2. *La relation de Chasles est vérifiée.*
3. *Si $f: [a,b] \to \mathbb{R}^r$ est continue, alors elle est intégrable.*
4. *Si $f: [a,b] \to \mathbb{R}^r$ est continument dérivable, alors*

$$\int_a^b f'(t)\,dt = f(b) - f(a).$$

La démonstration de cette proposition est laissée en exercice ; c'est une conséquence directe des définitions. Nous allons nous contenter de la remarque suivante. Si $e_1, \dots, e_r$ est une base de $\mathbb{R}^r$, alors pour chaque t on peut écrire

$$f(t) = \lambda_1(t)e_1 + \dots + \lambda_r(t)e_r.$$

En utilisant les deux propriétés énoncées dans le (1) de la proposition, on en déduit, si f est intégrable, que

$$\int_a^b f(t)\,dt = \int_a^b \lambda_1(t)e_1\,dt + \dots + \int_a^b \lambda_r(t)e_r\,dt$$
$$= \left(\int_a^b \lambda_1(t)\,dt \right) e_1 + \dots + \left(\int_a^b \lambda_r(t)\,dt \right) e_r.$$

Si maintenant on prend pour $e_1, \dots, e_r$ la base canonique, on retrouve la définition même de l'intégrale de f ; en d'autres termes, pour que le (1) de la proposition soit vrai, la seule définition possible est celle que nous avons donnée. Dans le même temps, nous observons que l'intégrale se comporte « comme prévu » dans toutes les bases, donc notre définition ne privilégie pas la base canonique.

Pour travailler avec les fonctions à valeurs vectorielles, il serait utile de montrer que l'on peut se ramener aux fonctions en escaliers, comme dans le cas des fonctions à valeurs dans $\mathbb{R}$. La définition de fonction « en escaliers » ne change pas : il s'agit toujours d'une fonction $\phi: [a,b] \to \mathbb{R}^r$ qui est constante sur $]a_k, a_{k+1}[$ pour une certaine subdivision (à ceci près que la valeur constante est bien sûr un vecteur).

Il y a une difficulté : écrire $\phi \le f$ n'a pas de sens pour les fonctions à valeurs dans $\mathbb{R}^r$. Pour cette raison, on a un résultat moins précis que le lemme 8.7, mais qui va suffire.

Proposition 8.28 – *Soit $f: [a,b] \to \mathbb{R}^r$ une fonction intégrable. Alors il existe une suite de fonctions en escaliers $(\phi_n)_{n\ge 0}$ à valeurs dans $\mathbb{R}^r$ telle que*

$$\int_a^b \phi_n(t)\,dt \xrightarrow[n\to\infty]{} \int_a^b f(t)\,dt. \qquad (1)$$

Si de plus f est continue, alors on a

$$\int_a^b \|f(t) - \phi_n(t)\|\,dt \xrightarrow[n\to\infty]{} 0, \qquad (2)$$

et enfin

$$\int_a^b \|\phi_n(t)\|\,dt \xrightarrow[n\to\infty]{} \int_a^b \|f(t)\|\,dt. \qquad (3)$$

Remarquons que l'hypothèse « f continue » est uniquement là pour nous simplifier la vie. En effet, pour que les relations (2) et (3) aient un sens, il faut que la fonction $t \mapsto \|f(t)\|$ soit intégrable, et même la fonction $t \mapsto \|f(t) + v\|$ pour tout vecteur constant v (ainsi en utilisant la relation de Chasles on voit que $t \mapsto \|f(t) - \phi_n(t)\|$ est également intégrable). Ceci est automatique lorsque f est continue (cf (3) de la proposition). Dans la démonstration vous verrez bien que la continuité n'est pas utilisée pour autre chose.

Démonstration. Pour chaque indice k on applique le lemme 8.7 à f_k, et on trouve une suite $(\phi_{k,n})$ de fonctions en escaliers telles que $\phi_{k,n} \le f_k$ et

$$\int_a^b \phi_{k,n}(t)\,dt \xrightarrow[n\to\infty]{} \int_a^b f_k(t)\,dt. \qquad (*)$$

Posons $\phi_n(t) = (\phi_{1,n}(t), \dots, \phi_{r,n}(t))$. Alors ϕ_n est en escaliers, à valeurs dans $\mathbb{R}^r$. La relation (1) est alors une conséquence immédiate des définitions.

Pour montrer (2), qui est plus difficile à établir, soit $e_1, \dots, e_r$ la base canonique de $\mathbb{R}^r$. Écrivons

$$f(t) - \phi_n(t) = (f_1(t) - \phi_{1,n}(t))\,e_1 + \dots + (f_r(t) - \phi_{r,n}(t))\,e_r.$$

D'après l'inégalité triangulaire pour les vecteurs, on a

$$\|f(t) - \phi_n(t)\| \le |f_1(t) - \phi_{1,n}(t)|\,\|e_1\| + \dots + |f_r(t) - \phi_{r,n}(t)|\,\|e_r\|.$$

Puisque $f_k \ge \phi_{k,n}$ et $\|e_k\| = 1$, on peut réécrire ceci

$$\|f(t) - \phi_n(t)\| \le (f_1(t) - \phi_{1,n}(t)) + \dots + (f_r(t) - \phi_{r,n}(t)),$$

d'où

$$\int_a^b \|f(t) - \phi_n(t)\|\,dt \le \int_a^b f_1(t)\,dt - \int_a^b \phi_{1,n}(t)\,dt + \cdots$$
$$+ \int_a^b f_r(t)\,dt - \int_a^b \phi_{r,n}(t)\,dt.$$

Ainsi la relation (*) entraîne bien la relation (2).

Pour la (3), on utilise tout simplement la « deuxième » inégalité triangulaire, celle qui affirme que $|\|a\| - \|b\|| \le \|a - b\|$ (cf lemme 3.31). Ceci donne

$$\left| \int_a^b (\|f(t)\| - \|\phi_n(t)\|)\,dt \right| \le \int_a^b |\|f(t)\| - \|\phi_n(t)\||\,dt$$
$$\le \int_a^b \|f(t) - \phi_n(t)\|\,dt.$$

(On a utilisé aussi l'inégalité triangulaire pour les intégrales de fonctions réelles.) On constate que (3) est une conséquence de (2). □

À l'aide de la proposition 8.28, nous pouvons montrer l'inégalité triangulaire pour les fonctions à valeurs dans $\mathbb{R}^r$. Il ne vous aura pas échappé qu'il nous a fallu plus d'efforts pour l'obtenir que dans le cas des fonctions réelles.

Proposition 8.29 – *Soit $f: [a,b] \to \mathbb{R}^r$ continue. Alors*

$$\left\| \int_a^b f(t)\,dt \right\| \le \int_a^b \|f(t)\|\,dt.$$

Démonstration. Supposons d'abord que f est en escaliers. Son intégrale est alors de la forme

$$\sum_k (a_{k+1} - a_k)\alpha_k,$$

où f est constante de valeur α_k sur $]a_k, a_{k+1}[$. L'inégalité triangulaire pour les vecteurs donne

$$\left\| \sum_k (a_{k+1} - a_k)\alpha_k \right\| \le \sum_k (a_{k+1} - a_k)\|\alpha_k\|.$$

Or le membre de droite n'est autre que l'intégrale de la fonction en escaliers $t \mapsto \|f(t)\|$. Donc l'inégalité triangulaire est vraie pour les intégrales de fonctions en escaliers.

Pour f continue, on prend une suite (ϕ_n) comme dans la proposition 8.28. Pour chaque n on a

$$\left\| \int_a^b \phi_n(t)\,dt \right\| \le \int_a^b \|\phi_n(t)\|\,dt,$$

puisque ϕ_n est en escaliers. Le membre de droite tend vers l'intégrale de $t \mapsto \|f(t)\|$ d'après le (3) de la proposition. Le membre de gauche tend vers

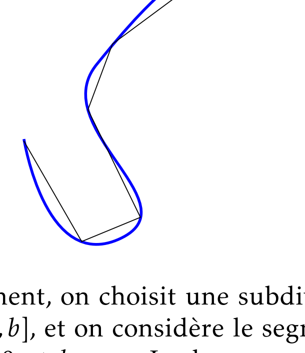
$$\left\| \int_a^b f(t)\,dt \right\|,$$

d'après le (1) de la proposition et le lemme 3.33. En passant à la limite sur n, on a donc l'inégalité annoncée. □

Nous allons utiliser ceci pour montrer que « le plus court chemin entre deux points, c'est la ligne droite ».

Une *courbe* est tout simplement une application $\gamma\colon [a,b] \to \mathbb{R}^r$. On utilise surtout le mot « courbe » dans les cas $r = 2$ (courbes planaires) ou $r = 3$ (courbes dans l'espace), mais leur étude peut se faire en général.

Comment définir la longueur d'une courbe γ ? Une idée naturelle est de chercher une approximation de la courbe par des segments de droite, comme dans la figure ci-dessous.



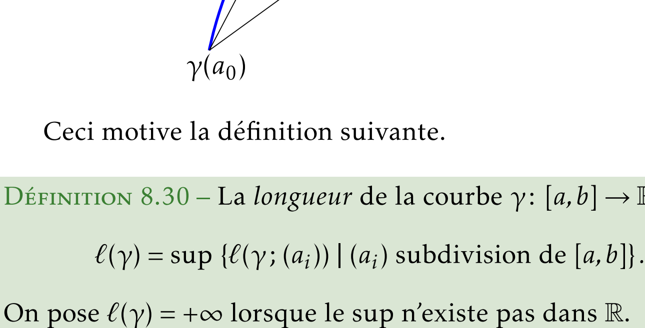
Plus précisément, on choisit une subdivision $a_0 = a < a_1 < \cdots < a_n = b$ de $[a,b]$, et on considère le segment qui joint $\gamma(a_k)$ à $\gamma(a_{k+1})$, pour $0 \le k < n$. La longueur de ce segment est $\|\gamma(a_{k+1}) - \gamma(a_k)\|$, donc une première approximation de la longueur de γ est

$$\ell(\gamma; a_0, a_1, \ldots, a_n) = \sum_{k=0}^{n-1} \|\gamma(a_{k+1}) - \gamma(a_k)\|.$$

Si on insère un nouveau point dans la subdivision, disons si on ajoute un point a_0' entre a_0 et a_1, alors on peut écrire

$$\begin{aligned}\|\gamma(a_1) - \gamma(a_0)\| &= \|\gamma(a_1) - \gamma(a_0') + \gamma(a_0') - \gamma(a_0)\| \\ &\le \|\gamma(a_1) - \gamma(a_0')\| + \|\gamma(a_0') - \gamma(a_0)\|,\end{aligned}$$

d'après l'inégalité triangulaire. En conséquence de cette observation, on a $\ell(\gamma; a_0, a_0', a_1, \ldots, a_n) \ge \ell(\gamma; a_0, a_1, \ldots, a_n)$: la longueur augmente à mesure que la subdivision devient plus fine.



Ceci motive la définition suivante.

Définition 8.30 – La *longueur* de la courbe $\gamma\colon [a,b] \to \mathbb{R}^r$ est

$$\ell(\gamma) = \sup\{\ell(\gamma; (a_i)) \mid (a_i) \text{ subdivision de } [a,b]\}.$$

On pose $\ell(\gamma) = +\infty$ lorsque le sup n'existe pas dans $\mathbb{R}$.

On dit parfois d'une courbe γ telle que $\ell(\gamma) < +\infty$ qu'elle est *rectifiable*.

Le résultat qui va rendre les choses calculables est le suivant :

Proposition 8.31 – *Soit $\gamma\colon [a,b] \to \mathbb{R}^r$ une courbe continument dérivable. Alors*

$$\ell(\gamma) = \int_a^b \|\gamma'(t)\|\,\mathrm{d}t.$$

Démonstration. Soit a une subdivision. On écrit

$$\begin{aligned}\|\gamma(a_{k+1}) - \gamma(a_k)\| &= \left\|\int_{a_k}^{a_{k+1}} \gamma'(t)\,\mathrm{d}t\right\| \\ &\le \int_{a_k}^{a_{k+1}} \|\gamma'(t)\|\,\mathrm{d}t, \qquad (*)\end{aligned}$$

en utilisant le théorème fondamental de l'analyse puis l'inégalité triangulaire. En faisant la somme sur tous les indices k, par la relation de Chasles il vient

$$\ell(\gamma, a) \le \int_a^b \|\gamma'(t)\|\,\mathrm{d}t,$$

et donc

$$\ell(\gamma) \le \int_a^b \|\gamma'(t)\|\,\mathrm{d}t.$$

Soit maintenant $\varepsilon > 0$. On va montrer qu'il existe une subdivision a telle que

$$\int_a^b \|\gamma'(t)\|\,\mathrm{d}t \le \ell(\gamma, a) + \varepsilon,$$

ce qui terminera la démonstration. Pour cela, examinons un cas simple dans lequel l'inégalité triangulaire est en fait une égalité : lorsque v est un vecteur constant, on a bien

$$\left\|\int_a^b v\,\mathrm{d}t\right\| = \|(b-a)v\| = (b-a)\|v\| = \int_a^b \|v\|\,\mathrm{d}t.$$

Prenons deux points a_k et a_{k+1} dans $[a,b]$, et appliquons cette dernière remarque avec $v = \gamma'(a_k)$. Intuitivement, l'idée est la suivante : lorsque a_k et a_{k+1} sont très proches, la fonction continue γ' ne varie pas beaucoup sur $[a_k, a_{k+1}]$, donc elle est presque constante, égale à $\gamma'(a_k)$; l'inégalité (*) ci-dessus est alors presque une égalité. Pour mettre en forme ceci, notons

$$\mathrm{M}_k = \sup\{\|\gamma'(t) - \gamma'(a_k)\| \mid t \in [a_k, a_{k+1}]\},$$

et écrivons :

$$\begin{aligned}\int_{a_k}^{a_{k+1}} \|\gamma'(t)\|\,\mathrm{d}t &= \int_{a_k}^{a_{k+1}} \|\gamma'(a_k) + (\gamma'(t) - \gamma'(a_k))\|\,\mathrm{d}t \\ &\le \int_{a_k}^{a_{k+1}} \|\gamma'(a_k)\|\,\mathrm{d}t + \int_{a_k}^{a_{k+1}} \|\gamma'(t) - \gamma'(a_k)\|\,\mathrm{d}t \\ &= \left\|\int_{a_k}^{a_{k+1}} \gamma'(a_k)\,\mathrm{d}t\right\| + \int_{a_k}^{a_{k+1}} \|\gamma'(t) - \gamma'(a_k)\|\,\mathrm{d}t \\ &\le \left\|\int_{a_k}^{a_{k+1}} \gamma'(a_k)\,\mathrm{d}t\right\| + \mathrm{M}_k(a_{k+1} - a_k) \\ &= \left\|\int_{a_k}^{a_{k+1}} [\gamma'(t) + (\gamma'(a_k) - \gamma'(t))]\,\mathrm{d}t\right\| + \mathrm{M}_k(a_{k+1} - a_k) \\ &\le \left\|\int_{a_k}^{a_{k+1}} \gamma'(t)\,\mathrm{d}t\right\| + \int_{a_k}^{a_{k+1}} \|\gamma'(a_k) - \gamma'(t)\|\,\mathrm{d}t \\ &\qquad + \mathrm{M}_k(a_{k+1} - a_k) \\ &\le \|\gamma(a_{k+1}) - \gamma(a_k)\| + 2\mathrm{M}_k(a_{k+1} - a_k).\end{aligned}$$

Puisque γ' est continue sur l'intervalle compact $[a,b]$, elle est uniformément continue d'après le théorème de Heine. Ainsi, il existe un $\delta > 0$ tel que $\|\gamma'(x) - \gamma'(y)\| < \frac{\varepsilon}{2(b-a)}$ dès que $|x - y| < \delta$. On peut alors choisir n un entier tel que $\frac{b-a}{n} < \delta$, et poser $a_k = a + k\left(\frac{b-a}{n}\right)$, de sorte que $|a_{k+1} - a_k| = \frac{b-a}{n} < \delta$ et donc $\mathrm{M}_k \le \frac{\varepsilon}{2(b-a)}$.

Si on fait la somme des inégalités

$$\int_{a_k}^{a_{k+1}} \|\gamma'(t)\|\,\mathrm{d}t \le \|\gamma(a_{k+1}) - \gamma(a_k)\| + \frac{\varepsilon}{n}$$

pour tous les indices k, on termine avec

$$\int_a^b \|\gamma'(t)\|\,\mathrm{d}t \le \ell(\gamma, a) + \varepsilon,$$

comme annoncé. □

Corollaire 8.32 – *Soient p et q deux points dans $\mathbb{R}^r$. Alors la longueur minimale d'une courbe joignant p et q est $\|q - p\|$, et ce minimum est atteint par la ligne droite.*

Démonstration. Pour montrer que la longueur d'une courbe γ joignant p à q est toujours supérieure à $\|q - p\|$, il suffit de considérer la subdivision ayant seulement deux points $a_0 = p$ et $a_1 = q$, puis de s'en rapporter à la définition.

Prenons maintenant la ligne droite, disons $\gamma\colon [0,1] \to \mathbb{R}^r$ définie par

$$\gamma(t) = (1-t)p + tq \quad (= p + t(q-p)).$$

(C'est bien un déplacement en ligne droite de p vers q.) La dérivée est $\gamma'(t) = q - p$, un vecteur constant, donc

$$\int_0^1 \|\gamma'(t)\|\,\mathrm{d}t = \|q - p\| = \ell(\gamma).$$

Ainsi dans le cas de la ligne droite le minimum est atteint. □

Exemple 8.33 (Circonférence d'un cercle) – Considérons, dans le plan complexe identifié à $\mathbb{R}^2$, le cercle de centre p et de rayon R. On peut le parcourir avec la courbe $\gamma\colon [0, 2\pi] \to \mathbb{C}$ définie par

$$\gamma(t) = p + \mathrm{R}e^{it},$$

comme expliqué dans le chapitre « L'exponentielle ». Calculons la longueur de cette courbe, qui est continument dérivable. On a $\gamma'(t) = \mathrm{R}ie^{it}$ d'où $\|\gamma'(t)\| = \mathrm{R}$. Ainsi

$$\ell(\gamma) = \int_0^{2\pi} \mathrm{R}\,\mathrm{d}t = 2\pi\mathrm{R}.$$

La circonférence d'un cercle de rayon R *est* 2πR, en particulier ça ne dépend pas du centre. Évidemment c'était la première définition historique du nombre π.

Il peut vous paraître surprenant qu'une courbe soit définie comme une *fonction* et pas un sous-ensemble de $\mathbb{R}^r$, et que la longueur d'une courbe $\gamma\colon [a,b] \to \mathbb{R}^r$ ne soit pas déterminée par l'image $\gamma([a,b])$. Par exemple, dans le cas du cercle, est-ce que la longueur change si l'on considère une courbe γ qui se déplace sur le cercle à une vitesse différente ? Et lorsque nous parlions ci-dessus de « la ligne droite » entre p et q, est-ce que l'on aurait pu considérer une autre paramétrisation que $t \mapsto (1-t)p + tq$, et trouver une autre longueur ?

La première réponse à ces questions est que la longueur dépend en effet en général de la fonction. L'exemple le plus bête est celui de la courbe $\gamma\colon [0, 4\pi] \to \mathbb{R}^2$ définie par $\gamma(t) = p + \mathrm{R}e^{it}$: en effet cette courbe fait « deux fois le tour » du cercle de centre p et de rayon R, et vous pouvez vérifier que sa longueur est 4πR. Alors que l'image $\gamma([0, 4\pi])$ est le cercle, qui peut être parcouru par une courbe de longueur 2πR comme on l'a vu. Ici les deux courbes sont très différentes dans leur comportement.

Cependant, voici un petit résultat qui exprime l'idée que des changements simples de paramétrisation ne vont pas changer la longueur.

Lemme 8.34 – *Soit $\gamma_1\colon [a,b] \to \mathbb{R}^r$ une courbe continument dérivable, et soit $u\colon [c,d] \to [a,b]$ une bijection, également supposée continument dérivable. Soit enfin $\gamma_2 = \gamma_1 \circ u$, qui est encore une courbe. Alors*

$$\ell(\gamma_1) = \ell(\gamma_2).$$

Démonstration. La fonction u est monotone d'après la proposition 4.23, on va supposer qu'elle est croissante (le cas décroissant est similaire), d'où $u(c) = a$, $u(d) = b$, et $u'(t) \ge 0$ pour tout $t \in [c,d]$.

On applique alors simplement le théorème du changement de variables :

$$\begin{aligned}\ell(\gamma_2) &= \int_c^d \|\gamma_2'(t)\|\,\mathrm{d}t \\ &= \int_c^d \|\gamma_1'(u(t))\,u'(t)\|\,\mathrm{d}t \\ &= \int_c^d \|\gamma_1'(u(t))\|\,u'(t)\,\mathrm{d}t \\ &= \int_a^b \|\gamma_1'(u)\|\,\mathrm{d}u \\ &= \ell(\gamma_1).\end{aligned}$$

□

Pour en revenir au cercle, la courbe $\gamma_1\colon [0, 2\pi] \to \mathbb{R}^2$ définie par $\gamma_1(t) = p + \mathrm{R}e^{it}$ et la courbe $\gamma_2\colon [0,1] \to \mathbb{R}^2$ définie par $\gamma_2(t) = p + \mathrm{R}e^{2i\pi t}$ sont liées comme ci-dessus, avec $u(t) = 2\pi t$, donc elles ont la même longueur (comme on le vérifie tout de suite).

voir les exercices 808, 809, 810 (questions sur les longueurs seulement; tracer les courbes, éventuellement à l'aide d'un ordinateur)

Nous allons conclure ce chapitre avec la démonstration du théorème de Taylor-Young dans sa forme générale (dans le chapitre sur les formules de Taylor nous avions une petite hypothèse restrictive). À l'aide des intégrales, c'est très facile.

Lemme 8.35 – *Soit* I *un intervalle contenant* 0 *et* $f : I \longrightarrow \mathbb{R}$ *une fonction intégrable. On suppose que* $f(t) = o(t^n)$ *pour un certain entier* n*. Si on pose*

$$F(x) = \int_0^x f(t)\,dt,$$

pour $x \in I$*, alors* $F(x) = o(x^{n+1})$.

Démonstration. Par hypothèse $f(t) = t^n h(t)$ avec $h(t) \longrightarrow 0$, donc si on se donne un $\varepsilon > 0$ il existe un $\delta > 0$ tel que $|h(t)| < \varepsilon$ pour $|t| < \delta$. Si on prend également $0 < x < \delta$ on a

$$\begin{aligned}\left|\int_0^x f(t)\,dt\right| &\le \int_0^x |f(t)|\,dt \\ &\le \int_0^x t^n \varepsilon\,dt \\ &= \varepsilon \left[\frac{t^{n+1}}{n+1}\right]_0^x = \varepsilon \frac{x^{n+1}}{n+1}.\end{aligned}$$

On traite de la même façon le cas $-\delta < x < 0$, et finalement on constate que si $|x| < \delta$, alors

$$\left|\frac{F(x)}{x^{n+1}}\right| \le \frac{\varepsilon}{n+1}.$$

C'est donc que

$$\frac{F(x)}{x^{n+1}} \longrightarrow 0$$

lorsque $x \longrightarrow 0$, comme on le souhaitait. □

On peut alors montrer facilement

Théorème 8.36 (Taylor-Young) – *Soit* f *une fonction dérivable* n *fois sur un intervalle* I *contenant* 0*. Alors on peut écrire*

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + o(x^n).$$

Démonstration. Par récurrence. Le cas $n = 1$ est donné par le lemme 6.7. Supposons le théorème vrai pour n, et soit f dérivable $n+1$ fois. Appliquons le résultat au rang n à la fonction f', qui est dérivable n fois :

$$f'(t) = f'(0) + (f')'(0)t + \cdots + \frac{(f')^{(n)}(0)}{n!}t^n + o(t^n).$$

En intégrant ceci entre 0 et x, et en utilisant le lemme précédent, on obtient la formule pour f au rang $n+1$. □

Chapitre 9

Fractions rationnelles

Le lecteur ayant assimilé la définition 2.17 peut prendre pour $\mathbb{K}$ n'importe quel corps.

Dans ce chapitre, la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$. L'étude des fractions rationnelles est un sujet algébrique, mais ce chapitre va surtout se concentrer sur l'application au calcul de certaines intégrales. Voilà pourquoi nous sommes encore dans la partie « Analyse ».

Définition 9.1 – Une *fraction rationnelle* F à coefficients dans $\mathbb{K}$ est un quotient de deux polynômes de $\mathbb{K}[X]$:

$$F = \frac{P}{Q},$$

avec $Q \neq 0$. L'ensemble des fractions rationnelles est noté $\mathbb{K}(X)$.

Les règles de calcul usuelles s'appliquent. Dans le jargon de la définition 2.17, on dira que $\mathbb{K}(X)$ est un corps. Des informations supplémentaires sont apportées dans l'encadré « Les corps de fractions ».

Les corps de fractions

La question désormais classique revient : comment donner une définition « complète » des fractions rationnelles, suffisamment complète pour qu'un ordinateur puisse faire des calculs ?
Il se trouve que le procédé est essentiellement identique, quasiment mot pour mot, à celui décrit dans l'encadré « Une définition de $\mathbb{Q}$ » dans le chapitre « Nombres » ; à ceci près qu'il faut remplacer $\mathbb{Z}$ par $\mathbb{K}[X]$. En fait cette construction porte un nom : on dit que $\mathbb{Q}$ est le *corps des fractions* de l'anneau $\mathbb{Z}$, et que $\mathbb{K}(X)$ est le corps des fractions de l'anneau $\mathbb{K}[X]$.
Pour faire cette construction à partir d'un anneau A quelconque, il y a une petite restriction technique : il faut pouvoir « simplifier » dans A. Plus précisément, on dit qu'un anneau A est *intègre* lorsque l'égalité $ab = 0$ avec $a, b \in A$ et $a \neq 0$, entraîne $b = 0$. C'est vrai pour $A = \mathbb{Z}$ et $A = \mathbb{K}[X]$, mais pas pour $A = M_n(\mathbb{K})$.
Vous pourrez montrer alors facilement, en copiant ce que l'on a fait pour $\mathbb{Q}$, qu'à partir de n'importe quel anneau A qui est commutatif et intègre, on peut contruire un corps, noté $Fr(A)$ et appelé le corps des fractions de A.
Il y aura peu d'exemples en première année, hormis $\mathbb{Z}$ et $\mathbb{K}[X]$. On peut en donner un tout de même en considérant $A = \mathbb{Z}[\sqrt{2}]$, qui par définition est l'ensemble des nombres de la forme $a + b\sqrt{2}$ avec $a, b \in \mathbb{Z}$. Vous montrerez que cet anneau est commutatif et intègre, et que son corps des fractions est $\mathbb{Q}[\sqrt{2}]$, c'est-à-dire l'ensemble des nombres de la forme $a + b\sqrt{2}$ avec cette fois $a, b \in \mathbb{Q}$.

Le nombre $\deg P - \deg Q$ est appelé le *degré* de F, noté $\deg F$. Il faut vérifier qu'il est bien défini, puisque la paire (P, Q) n'est pas unique : en fait on a

$$\frac{P}{Q} = \frac{P'}{Q'} \iff PQ' = P'Q,$$

et ainsi $\deg P + \deg Q' = \deg P' + \deg Q$, d'où $\deg P - \deg Q = \deg P' - \deg Q'$, comme souhaité.

Notre objectif est de mettre les fractions rationnelles sous une forme particulière, qui facilite le calcul des primitives lorsque $\mathbb{K} = \mathbb{R}$. Tout part de l'observation suivante :

Lemme 9.2 – *Soit $F = \frac{P}{Q}$ une fraction rationnelle. Supposons que l'on ait une factorisation $Q = AB$ avec $\mathrm{pgcd}(A, B) = 1$. Alors il existe des polynômes P_1 et P_2 tels que*

$$F = \frac{P_1}{A} + \frac{P_2}{B}.$$

Démonstration. On utilise le théorème de Bézout (11.22), qui nous donne l'existence de U et V tels que $AU + BV = 1$. Par suite

$$\frac{1}{AB} = \frac{AU + BV}{AB} = \frac{U}{B} + \frac{V}{A}.$$

En multipliant par P, il vient

$$F = \frac{P}{AB} = \frac{PV}{A} + \frac{PU}{B},$$

d'où le résultat avec $P_1 = PV$ et $P_2 = PU$. □

On sait donc « couper » une fraction rationnelle en deux lorsque son dénominateur se factorise en deux termes premiers entre eux. À l'opposé, on a une autre remarque calculatoire :

Lemme 9.3 – *Soient P et Q deux polynômes, soit α un entier, et soit*

$$F = \frac{P}{Q^\alpha}.$$

Alors il existe des polynômes P_j, pour $0 \leq j \leq \alpha$, tels que

$$F = P_0 + \frac{P_1}{Q} + \frac{P_2}{Q^2} + \cdots + \frac{P_\alpha}{Q^\alpha},$$

et tels que $\deg P_j < \deg Q$ pour $j \geq 1$.

Attention, il n'y a pas de condition sur le degré de P_0.

Démonstration. Faisons une division euclidienne

$$P = AQ + R,$$

avec $\deg R < \deg Q$. Alors

$$\frac{P}{Q^\alpha} = \frac{AQ + R}{Q^\alpha} = \frac{A}{Q^{\alpha-1}} + \frac{R}{Q^\alpha}.$$

On peut alors poser $P_\alpha = R$ et faire une récurrence sur α. □

On en déduit le théorème principal de ce chapitre.

Théorème 9.4 (Décomposition en éléments simples) – *Soit F une fraction rationnelle. Écrivons*

$$F = \frac{P}{Q} = \frac{P}{Q_1^{\alpha_1} Q_2^{\alpha_2} \cdots Q_n^{\alpha_n}},$$

avec chaque Q_i irréductible, et avec $\mathrm{pgcd}(Q_i, Q_j) = 1$ pour $i \neq j$. Alors il existe des polynômes $P_{i,j}$ pour $1 \leq i \leq n$ et $1 \leq j \leq \alpha_i$, et un polynôme P_0, tels que

$$F = P_0 + \sum_{i=1}^{n} \sum_{j=1}^{\alpha_i} \frac{P_{i,j}}{Q_i^j},$$

avec $\deg P_{i,j} < \deg Q_i$.

De plus, si $\deg P < \deg Q$, alors $P_0 = 0$.

Cet énoncé peut paraître compliqué, mais nous verrons qu'il est facile à mettre en pratique. L'essentiel est que chaque terme

$$\frac{P_{i,j}}{Q_i^j},$$

que l'on appelle « élément simple », possède une primitive assez facile à calculer (lorsque $\mathbb{K} = \mathbb{R}$). On va donc être capable de calculer une primitive de n'importe quelle fraction rationnelle F.

Démonstration. C'est la combinaison des deux lemmes précédents. Le deuxième lemme traite le cas $n = 1$. Dans le cas général, on pose $A = Q_1^{\alpha_1} \cdots Q_{n-1}^{\alpha_{n-1}}$ et $B = Q_n^{\alpha_n}$, de sorte que $\mathrm{pgcd}(A, B) = 1$. Le premier lemme donne alors

$$F = \frac{P_1}{A} + \frac{P_2}{B},$$

et l'on peut faire une récurrence sur n.

Reste à montrer le « de plus ». Notons une chose simple : si F et G sont deux fractions rationnelles, on a

$$\deg(F + G) \leq \max(\deg F, \deg G).$$

Dans notre théorème, si $\deg F < 0$, on en déduit que $\deg P_0 < 0$. Or P_0 est un polynôme, donc ceci impose $P_0 = 0$. □

Avant de donner des exemples de calculs, donnons des versions de ce théorème spécialisées à $\mathbb{R}$ et $\mathbb{C}$.

Corollaire 9.5 (Éléments simples sur $\mathbb{C}$) – *Soit*

$$F = \frac{P}{(X - x_1)^{\alpha_1} \cdots (X - x_n)^{\alpha_n}} \in \mathbb{C}(X),$$

où les nombres x_i sont distincts. Alors on peut écrire

$$F = P_0 + \sum_{i=1}^{n} \sum_{j=1}^{\alpha_i} \frac{\lambda_{i,j}}{(X - x_i)^j},$$

où P_0 est un polynôme et $\lambda_{i,j} \in \mathbb{C}$. De plus si $\deg F < 0$ alors $P_0 = 0$.

exercice : montrer que les $\lambda_{i,j}$ sont en fait uniques

Exemple 9.6 – Prenons

$$F = \frac{X^3 - X}{X^2 + 1}.$$

Le premier réflexe est de faire une division euclidienne du numérateur par le dénominateur. Ici on a $X^3 - X = X(X^2 + 1) - 2X$, de sorte que

$$F = X - \frac{2X}{X^2 + 1}.$$

On va maintenant appliquer le théorème au dernier terme à droite. Puisque $X^2 + 1 = (X - i)(X + i)$, on sait qu'il existe λ et μ tels que

$$\frac{2X}{X^2 + 1} = \frac{\lambda}{X - i} + \frac{\mu}{X + i}.$$

Le plus simple pour trouver les valeurs de λ et μ n'est pas de procéder comme dans la démonstration du théorème, mais simplement d'identifier les numérateurs :

$$\frac{\lambda}{X - i} + \frac{\mu}{X + i} = \frac{(\lambda + \mu)X + (\lambda - \mu)i}{X^2 + 1} = \frac{2X}{X^2 + 1}.$$

On a donc $\mu - \lambda = 0$ et $\lambda + \mu = 2$, donc $\lambda = \mu = 1$. Finalement

$$F = X - \frac{1}{X - i} - \frac{1}{X + i}.$$

Voici maintenant la version sur $\mathbb{R}$, qui est celle qui nous intéresse le plus pour calculer des primitives.

Corollaire 9.7 (Éléments simples sur $\mathbb{R}$) – *Soit*

$$F = \frac{P}{(X - x_1)^{\alpha_1} \cdots (X - x_n)^{\alpha_n} Q_1^{\beta_1} \cdots Q_m^{\beta_m}} \in \mathbb{R}(X),$$

où les nombres x_i sont distincts, chaque polynôme Q_i est de degré 2 sans racine réelle, et $\mathrm{pgcd}(Q_i, Q_j) = 1$ si $i \neq j$. Alors on peut écrire

$$F = P_0 + \sum_{i=1}^{n} \sum_{j=1}^{\alpha_i} \frac{\lambda_{i,j}}{(X - x_i)^j} + \sum_{i=1}^{m} \sum_{j=1}^{\beta_i} \frac{\mu_{i,j} X + \gamma_{i,j}}{Q_i^j},$$

où $P_0 \in \mathbb{R}[X]$, et $\lambda_{i,j}, \mu_{i,j}, \gamma_{i,j} \in \mathbb{R}$. De plus si $\deg F < 0$ alors $P_0 = 0$.

Exemple 9.8 – Si on reprend l'exemple précédent, c'est-à-dire

$$F = \frac{X^3 - X}{X^2 + 1},$$

alors on a vu qu'après une simple division euclidienne on a

$$F = X - \frac{2X}{X^2 + 1}.$$

Sur $\mathbb{R}$, on s'arrête là : on ne peut pas factoriser plus le dénominateur $X^2 + 1$, et l'expression obtenue est bien de la forme annoncée dans le corollaire. Notons d'ailleurs que l'on a assez travaillé pour calculer une primitive, en effet

$$\int_a^b F(x)\,dx = \left[\frac{x^2}{2} - \ln(x^2 + 1)\right]_a^b.$$

Voyons un exemple beaucoup plus compliqué. Prenons

$$F = \frac{1}{X^5 - 2X^4 + 6X^3 - 12X^2 + 9X - 18}.$$

Il faut d'abord factoriser le dénominateur. Nous avons de la chance, puisque 2 est une racine « évidente ». En faisant une division euclidienne, on obtient $(X - 2)(X^4 + 6X^2 + 9)$. De plus $X^4 + 6X^2 + 9 = (X^2 + 3)^2$, de sorte que

$$F = \frac{1}{(X - 2)(X^2 + 3)^2}.$$

Le corollaire nous annonce donc que nous pouvons écrire

$$F = \frac{\lambda}{X - 2} + \frac{\mu_1 X + \gamma_1}{X^2 + 3} + \frac{\mu_2 X + \gamma_2}{(X^2 + 3)^2}. \qquad (*)$$

Il est possible, bien que très long, de mettre le membre de droite au même dénominateur. On obtient au numérateur

$$(\lambda + \mu_1)X^4 + (-2\mu_1 + \gamma_1)X^3 + (6\lambda + 3\mu_1 - 2\gamma_1 + \mu_2)X^2 + (-6\mu_1 + 3\gamma_1 - 2\mu_2 + \gamma_2)X + (9\lambda - 6\gamma_1 - 2\gamma_2). \qquad (**)$$

En écrivant que ce numérateur doit valoir 1, on obtient alors un système à 5 inconnues et 5 équations, dont on sait qu'il possède des solutions, et on peut les trouver par les techniques habituelles. À l'aide d'un ordinateur, c'est une méthode très efficace.

Bien sûr pour finir le calcul « à la main », on peut utiliser des astuces. Voici une compilation des plus connues. Multiplions l'équation (*) par le polynôme $X - 2$:

$$(X-2)F = \frac{1}{(X^2 + 3)^2} = \lambda + \frac{(\mu_1 X + \gamma_1)(X - 2)}{X^2 + 3} + \frac{(\mu_2 X + \gamma_2)(X - 2)}{(X^2 + 3)^2}.$$

On peut voir ça comme une égalité de fonctions, et on va regarder la valeur en $X = 2$: il reste simplement $\lambda = \frac{1}{49}$. On a déjà trouvé λ !

Maintenant multiplions (*) par $(X^2+3)^2$; nous obtenons

$$\frac{1}{X - 2} = \frac{1}{49}\frac{(X^2 + 3)^2}{X - 2} + (\mu_1 X + \gamma_1)(X^2 + 3) + (\mu_2 X + \gamma_2).$$

On évalue en $X = i\sqrt{3}$ (pour que $X^2 + 3 = 0$), et il reste simplement

$$\frac{1}{i\sqrt{3} - 2} = \mu_2 i\sqrt{3} + \gamma_2.$$

Pour trouver les parties réelle et imaginaire du membre de gauche on écrit bien sûr

$$\frac{1}{i\sqrt{3} - 2} = \frac{-2 - i\sqrt{3}}{7},$$

d'où $\gamma_2 = -\frac{2}{7}$ et $\mu_2 = -\frac{1}{7}$.

Ensuite, on peut choisir d'évaluer (*) en $X = 0$, parce que c'est relativement facile : on obtient

$$-\frac{1}{18} = \frac{1}{3}\gamma_1 - \frac{37}{882},$$

d'où $\gamma_1 = -\frac{2}{49}$.

Regardons le numérateur (**) ci-dessus. Il commence par le terme $(\lambda + \mu_1)X^4$, ce que l'on peut vérifier de tête très vite sans tout mettre au même dénominateur. On a donc $\lambda + \mu_1 = 0$ d'où $\mu_1 = -\lambda = -\frac{1}{49}$.

Finalement

$$F = \frac{1}{49}\left(\frac{1}{X - 2} - \frac{X + 2}{X^2 + 3} - \frac{7X + 14}{(X^2 + 3)^2}\right).$$

Ce calcul était relativement compliqué. Pourrait-on confier une partie de la tâche à un ordinateur ? La seule chose qui n'était pas purement mécanique dans le raisonnement ci-dessus était... de trouver que 2 était une racine du dénominateur. Ceci étant, un ordinateur peut (au pire) mettre les deux membres de (*) au même dénominateur et résoudre un système, et ce, en une fraction de seconde. Retenons :

Proposition 9.9 – *Lorsque l'on sait factoriser complètement le dénominateur d'une fraction rationnelle, trouver la décomposition en éléments simples est une procédure automatisable, que l'on peut confier à une machine.*

Par contre, vous aurez peut-être besoin de savoir décomposer une fraction à la main, pour les besoins d'un concours ou d'un examen ! Dans l'absolu, il est utile de savoir traiter les cas très simples, mais il est absurde de devenir expert.

voir l'exercice 824 (les décompositions seulement)

Nous fixons désormais $\mathbb{K} = \mathbb{R}$. La décomposition en éléments simples est censée nous aider à calculer les primitives, et pour mettre cela en oeuvre il faut savoir intégrer chacun des termes apparaissant dans le corollaire 9.7.

Les plus simples sont bien sûr

$$\int_p^q \frac{\mathrm{d}x}{(x-x_0)^\alpha} = \left[\frac{(x-x_0)^{-\alpha+1}}{-\alpha+1} \right]_p^q$$

pour $\alpha > 1$ et

$$\int_p^q \frac{\mathrm{d}x}{x-x_0} = [\ \ln|x-x_0|\ ]_p^q$$

(ne pas oublier la valeur absolue). Il nous reste donc à traiter

$$\int_p^q \frac{\lambda x + \mu}{(ax^2+bx+c)^\alpha}\,\mathrm{d}x .$$

Il y a toute une série d'étapes pour y arriver. Le premier réflexe est de « faire apparaître la dérivée du dénominateur au numérateur ». En effet on sait calculer

$$\int_p^q \frac{2ax+b}{(ax^2+bx+c)^\alpha}\,\mathrm{d}x = \left[\frac{(ax^2+bx+c)^{\alpha+1}}{\alpha+1} \right]_p^q$$

pour $\alpha > 1$ et

$$\int_p^q \frac{2ax+b}{ax^2+bx+c}\,\mathrm{d}x = \left[\ \ln|ax^2+bx+c|\ \right]_p^q .$$

Donc on va se débrouiller pour faire apparaître $2ax+b$.

Exemple 9.10 – À la fin de l'exemple 9.8 nous avions le terme

$$\frac{x+2}{x^2+3} .$$

On écrit alors

$$\frac{x+2}{x^2+3} = \frac{1}{2}\left(\frac{2x}{x^2+3} + \frac{4}{x^2+3} \right),$$

de sorte que l'on sait intégrer une partie au moins de l'expression :

$$\begin{aligned} \int_p^q \frac{x+2}{x^2+3}\,\mathrm{d}x &= \frac{1}{2}\int_p^q \frac{2x}{x^2+3}\,\mathrm{d}x + 2\int_p^q \frac{1}{x^2+3}\,\mathrm{d}x \\ &= \frac{1}{2}\left[\ln(x^2+3) \right]_p^q + 2\int_p^q \frac{\mathrm{d}x}{x^2+3} . \end{aligned}$$

De la même manière si nous avons à intégrer l'expression

$$\frac{5x-3}{x^2+x+1},$$

la première chose à faire est d'écrire

$$\frac{5x-3}{x^2+x+1} = \frac{5}{2}\left(\frac{2x+1}{x^2+x+1} \right) - \frac{11}{2(x^2+x+1)} .$$

On a alors

$$\int_p^q \frac{5x-3}{x^2+x+1}\,\mathrm{d}x = \frac{5}{2}\left[\ \ln(x^2+x+1)\ \right]_p^q - \frac{11}{2}\int_p^q \frac{\mathrm{d}x}{x^2+x+1} .$$

Retournons au cas général. Nous sommes ramenés à calculer les intégrales de la forme

$$\int_p^q \frac{\mathrm{d}x}{(ax^2+bx+c)^n},$$

avec $b^2 - 4ac < 0$ puisqu'on suppose que le dénominateur ne s'annule pas. On en connaît une :

$$\int_p^q \frac{\mathrm{d}x}{x^2+1} = [\ \arctan(x)\ ]_p^q .$$

Nous allons voir que l'on peut toujours se ramener à ce cas-là (ce qui explique la présence abondante de la fonction arctangente dans toutes les questions de primitives). La prochaine étape est de faire un changement de variables pour mettre le dénominateur sous la forme $u(x)^2+1$.

Exemple 9.11 – Reprenons les exemples ci-dessus. On a

$$\frac{1}{x^2+3} = \frac{1}{3(\frac{x^2}{3}+1)} = \frac{1}{3((\frac{x}{\sqrt{3}})^2+1)} = \frac{1}{3(u(x)^2+1)}$$

avec $u(x) = \frac{x}{\sqrt{3}}$. On a $u'(x) = \frac{1}{\sqrt{3}}$ et donc

$$\begin{aligned} \int_p^q \frac{\mathrm{d}x}{x^2+3} &= \frac{\sqrt{3}}{3}\int_p^q \frac{\frac{1}{\sqrt{3}}\,\mathrm{d}x}{u(x)^2+1} = \frac{\sqrt{3}}{3}\int_p^q \frac{u'(x)\,\mathrm{d}x}{u(x)^2+1} \\ &= \frac{\sqrt{3}}{3}\int_{p/\sqrt{3}}^{q/\sqrt{3}} \frac{\mathrm{d}u}{u^2+1} = \frac{\sqrt{3}}{3}[\ \arctan(u)\ ]_{p/\sqrt{3}}^{q/\sqrt{3}} . \end{aligned}$$

Plutôt que de refaire ce changement de variables à chaque fois, on peut décider de mémoriser qu'une primitive de $x \mapsto \frac{1}{x^2+a^2}$ est $x \mapsto \frac{1}{a}\arctan(\frac{x}{a})$ (tout dépend de la quantité de primitives que l'on souhaite garder en tête). Pour $a = \sqrt{3}$ on retrouve le résultat ci-dessus évidemment (noter que $\frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$). Cependant, il est parfois préférable de refaire le changement de variables, notamment lorsque le dénominateur est élevé à une puissance (voir plus bas).

Essayons maintenant x^2+x+1 au dénominateur. La présence du terme de degré 1 nous force à une étape préliminaire :

$$x^2+x+1 = \left(x+\frac{1}{2}\right)^2 - \frac{1}{4} + 1 = t(x)^2 + \frac{3}{4}$$

avec $t(x) = x + \frac{1}{2}$. On peut faire un premier changement de variables :

$$\int_p^q \frac{\mathrm{d}x}{x^2+x+1} = \int_p^q \frac{t'(x)\,\mathrm{d}x}{t(x)^2+\frac{3}{4}} = \int_{p+\frac{1}{2}}^{q+\frac{1}{2}} \frac{\mathrm{d}t}{t^2+\frac{3}{4}} .$$

Ou bien on change encore de variables, ou bien on utilise la primitive que l'on connaît, pour en arriver à :

$$\int_{p+\frac{1}{2}}^{q+\frac{1}{2}} \frac{\mathrm{d}t}{t^2+\frac{3}{4}} = \left[\frac{2}{\sqrt{3}}\arctan\left(\frac{2t}{\sqrt{3}}\right) \right]_{p+\frac{1}{2}}^{q+\frac{1}{2}} .$$

En utilisant ce genre de changements de variables, on en arrive finalement toujours à calculer une intégrale de la forme

$$\mathrm{I}_n = \int_p^q \frac{\mathrm{d}x}{(x^2+1)^n} .$$

Nous savons faire pour $n = 1$, et la dernière chose que nous devons apprendre est le calcul de I_n pour tout entier n.

C'est le moment d'utiliser la technique, que vous avez vue au lycée, de *l'intégration par parties*. Le principe est très simple : de la formule

$$(fg)' = f'g + fg',$$

on tire $f'g = (fg)' - fg'$ et donc

$$\int_p^q f'(x)g(x)\,\mathrm{d}x = [f(x)g(x)]_p^q - \int_p^q f(x)g'(x)\,\mathrm{d}x .$$

(Cette formule est valable dès que f et g sont continument dérivables.)

Exemple 9.12 – Pour trouver une primitive de $x \mapsto \ln(x)$, on peut faire une intégration par parties avec $f(x) = x$ et $g(x) = \ln(x)$, d'où

$$\begin{aligned} \int_p^q \ln(x)\,\mathrm{d}x &= [\ x\ln(x)\ ]_p^q - \int_p^q \mathrm{d}x \\ &= [x\ln(x) - x]_p^q . \end{aligned}$$

Une primitive est donc $x \mapsto x\ln(x) - x$.

Pour en revenir au calcul de I_n, on a la formule de récurrence suivante.

Lemme 9.13 – *Pour $n \geq 1$ on a*

$$\mathrm{I}_{n+1} = \frac{2n-1}{2n}\mathrm{I}_n + \frac{1}{2n}\left[\frac{x}{(x^2+1)^n} \right]_p^q .$$

Plutôt que d'apprendre cette relation par coeur, il vaut mieux retenir que la démonstration, que voici, s'appuie sur une intégration par parties.

Démonstration. On prend $f(x) = 1$ et $g(x) = \frac{1}{(x^2+1)^n}$, d'où

$$\mathrm{I}_n = \left[\frac{x}{(x^2+1)^n} \right]_p^q + 2n\int_p^q \frac{x^2\,\mathrm{d}x}{(x^2+1)^{n+1}} .$$

En écrivant $x^2 = x^2+1-1$, on constate que la dernière intégrale à droite est $\mathrm{I}_n - \mathrm{I}_{n+1}$. On en tire le résultat en arrangeant les termes. □

Exemple 9.14 – Pour calculer

$$\int_p^q \frac{\mathrm{d}x}{(x^2+3)^2}$$

on commence par poser $u(x) = \frac{x}{\sqrt{3}}$ comme dans l'exemple précédent, de sorte que

$$\int_p^q \frac{\mathrm{d}x}{(x^2+3)^2} = \frac{\sqrt{3}}{9}\int_{p'}^{q'} \frac{\mathrm{d}u}{(u^2+1)^2}$$

avec $p' = \frac{p}{\sqrt{3}}$ et $q' = \frac{q}{\sqrt{3}}$. On fait une intégration par parties :

$$\begin{aligned} \int_{p'}^{q'} \frac{\mathrm{d}u}{u^2+1} &= \left[\frac{u}{u^2+1} \right]_{p'}^{q'} + 2\int_{p'}^{q'} \frac{u^2+1-1}{(u^2+1)^2}\,\mathrm{d}u \\ &= \left[\frac{u}{u^2+1} + 2\arctan(u) \right]_{p'}^{q'} - 2\int_{p'}^{q'} \frac{\mathrm{d}u}{(u^2+1)^2}, \end{aligned}$$

d'où

$$\int_{p'}^{q'} \frac{\mathrm{d}u}{(u^2+1)^2} = \frac{1}{2}\left[\frac{u}{(u^2+1)^2} + \arctan(u) \right]_{p'}^{q'} .$$

Finalement

$$\int_p^q \frac{\mathrm{d}x}{(x^2+3)^2} = \frac{\sqrt{3}}{18}\left[\frac{u}{u^2+1} + \arctan(u) \right]_{p/\sqrt{3}}^{q/\sqrt{3}} .$$

Ces calculs sont difficiles. Cependant, nous pouvons à nouveau remarquer que rien n'empêche un ordinateur de les faire pour nous : toutes les étapes sont parfaitement mécaniques.

Proposition 9.15 – *Lorsque l'on sait factoriser complètement le dénominateur d'une fraction rationnelle, trouver une primitive est une procédure automatisable, que l'on peut confier à une machine.*

Là encore, renseignez-vous pour savoir si l'on attend de vous, à l'occasion d'un concours ou d'un examen, que vous sachiez calculer ces primitives « à la main ». C'est très probable ! En tout cas il est souhaitable de savoir traiter quelques exemples. *voir l'exercice 824*

Nous allons apprendre à calculer des intégrales d'un certain type précis, à savoir de la forme

$$\int_p^q \mathrm{F}(\cos(\theta), \sin(\theta))\, \mathrm{d}\theta.$$

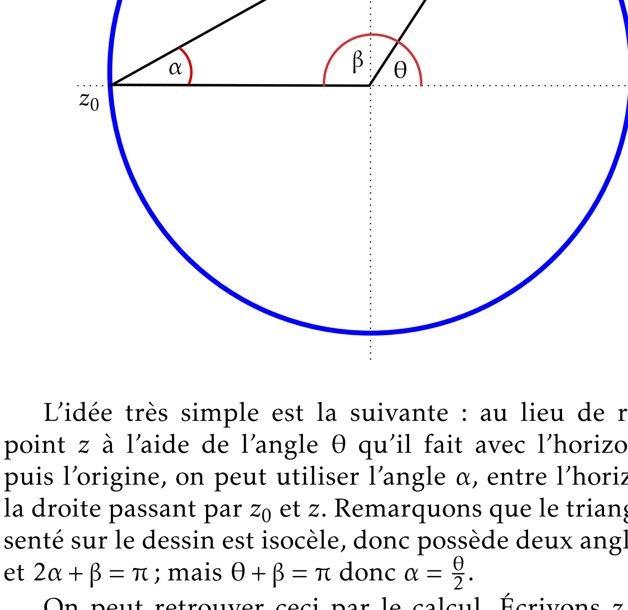
Ici on suppose que l'on a une fonction $(x,y) \mapsto \mathrm{F}(x,y)$, à deux variables, définie au moins aux points $(x,y) = (\cos(\theta), \sin(\theta))$ pour $p \leq \theta \leq q$. L'ensemble de ces points est un arc de cercle, et en faisant un peu de géométrie nous allons trouver un changement de variables judicieux. En particulier, l'intégrale ci-dessus va se ramener à une intégrale de fraction rationnelle lorsque $\mathrm{F}(x,y)$ est une expression utilisant seulement des opérations arithmétiques (addition, soustraction, multiplication et division), disons par exemple

$$\mathrm{F}(\cos(\theta), \sin(\theta)) = \frac{3\cos(\theta)^2 \sin(\theta) - 1}{\sin(\theta)^3 + \cos(\theta)}.$$

On dit parfois alors que $\mathrm{F}(\cos(\theta), \sin(\theta))$ est une « fraction rationnelle trigonométrique ».

Pour les impatients, il est très simple de résumer la méthode : utiliser le changement de variables $t(\theta) = \tan(\frac{\theta}{2})$. Vous pouvez de suite aller voir l'exemple 9.17 ci-dessous, en prenant connaissance des équations (**) au passage. Mais il est instructif de prendre le temps de comprendre pourquoi cette astuce fonctionne.

Considérons le dessin ci-dessous. On identifie le plan $\mathbb{R}^2$ avec $\mathbb{C}$, et sur le dessin on a placé le point $z_0 = -1$ ainsi qu'un point $z = x + iy = \cos(\theta) + i\sin(\theta)$ sur le cercle unité.



L'idée très simple est la suivante : au lieu de repérer le point z à l'aide de l'angle θ qu'il fait avec l'horizontale depuis l'origine, on peut utiliser l'angle α, entre l'horizontale et la droite passant par z_0 et z. Remarquons que le triangle représenté sur le dessin est isocèle, donc possède deux angles égaux, et $2\alpha + \beta = \pi$; mais $\theta + \beta = \pi$ donc $\alpha = \frac{\theta}{2}$.

On peut retrouver ceci par le calcul. Écrivons $z = x + iy$, alors le vecteur $u = z - z_0$ est simplement $u = z + 1 = x + 1 + iy$. Par définition de α, on a aussi $u = re^{i\alpha} = r\cos(\alpha) + ri\sin(\alpha)$ pour un certain $r > 0$, donc $\tan(\alpha) = \frac{y}{x+1}$. Puisque z est de module 1, on peut également écrire $x = \cos(\theta)$ et $y = \sin(\theta)$, de sorte que

$$\tan(\alpha) = \frac{\sin(\theta)}{\cos(\theta)+1} = \frac{2\cos(\frac{\theta}{2})\sin(\frac{\theta}{2})}{2\cos(\frac{\theta}{2})^2 - 1 + 1} = \frac{\sin(\frac{\theta}{2})}{\cos(\frac{\theta}{2})} = \tan\left(\frac{\theta}{2}\right).$$

On retrouve bien $\alpha = \frac{\theta}{2}$, si l'on prend θ dans l'intervalle $]-\pi, \pi[$ et α dans $]-\frac{\pi}{2}, \frac{\pi}{2}[$.

Nous venons d'utiliser quelques formules de trigonométrie bien connues. C'est en faisant un petit effort de calcul supplémentaire que l'on va comprendre l'intérêt d'utiliser α au lieu de θ. Posons $t = \tan(\frac{\theta}{2}) = \frac{y}{x+1}$, et calculons

$$1 + t^2 = \frac{y^2 + x^2 + 2x + 1}{(x+1)^2} = \frac{2}{x+1},$$

puisque $x^2 + y^2 = 1$. On en tire

$$x + 1 = \frac{2}{1+t^2}, \tag{*}$$

d'où

$$x = \cos(\theta) = \frac{1-t^2}{1+t^2} \quad \text{et} \quad y = \sin(\theta) = \frac{2t}{1+t^2}. \tag{**}$$

(La deuxième en multipliant (*) par t, puisque $y = t(x+1)$.) Ces relations sont très intéressantes, puisqu'elles nous poussent à utiliser en réalité t, et non pas α lui-même, pour repérer le point z sur le cercle : l'avantage est alors que les coordonnées x et y de z sont des fractions rationnelles en t. Par contraste, lorsque l'on exprime x et y en fonction de θ, on fait appel aux fonctions cosinus et sinus, qui sont bien plus compliquées (la différence va devenir très claire dans le calcul des primitives, ci-dessous).

Pour résumer, nous avons montré la chose suivante.

PROPOSITION 9.16 – *Soit $z = x + iy$ un nombre complexe ; on suppose $z \neq -1$. Alors z est de module 1 si et seulement s'il existe un nombre réel t tel que*

$$x = \frac{1-t^2}{1+t^2} \quad \text{et} \quad y = \frac{2t}{1+t^2}.$$

Dans ce cas, on a

$$t = \frac{y}{x+1},$$

donc en particulier t est uniquement déterminé par z.

Triplets pythagoriciens

La proposition 9.16 établit une bijection entre l'ensemble $\mathbb{R}$ d'une part, et l'ensemble des points sur le cercle unité (sauf $(-1,0)$) d'autre part. Cette bijection étant donnée par la formule explicite

$$t \mapsto \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right),$$

on constate qu'elle possède la propriété remarquable d'établir également une bijection entre $\mathbb{Q}$ et l'ensemble des points (x,y) sur le cercle tels que $x \in \mathbb{Q}$ et $y \in \mathbb{Q}$ (et $(x,y) \neq (-1,0)$).

Voici une application célèbre. Un *triplet pythagoricien* est donné par trois nombres entiers (a,b,c) tels que $a^2 + b^2 = c^2$. Grâce au théorème de Pythagore, on peut interpréter un tel triplet comme donnant les longueurs (entières !) d'un triangle rectangle. Par exemple $(3,4,5)$ est un triplet pythagoricien.

Connaissant un triplet (a,b,c), on peut en fabriquer une infinité en multipliant par un même nombre n, c'est-à-dire en considérant (na, nb, nc), mais les triangles correspondants ont le même aspect. Peut-on construire une infinité de triplets pythagoriciens, y compris en considérant comme identiques deux triplets « proportionnels » ? Sans parler d'infinité, peut-on déjà en construire beaucoup ? Combien pouvez-vous en citer ?

On commence par associer à tout triplet (a,b,c) la paire (x,y) avec $x = \frac{a}{c}$ et $y = \frac{b}{c}$. On a alors $x^2 + y^2 = 1$, c'est-à-dire que (x,y) est sur le cercle unité, et $x \in \mathbb{Q}$, $y \in \mathbb{Q}$. Vous montrerez à titre d'exercice que si un autre triplet (a',b',c') donne la paire (x',y'), alors $(x',y') = (x,y)$ si et seulement si (a',b',c') et (a,b,c) sont proportionnels. De plus, à partir de (x,y) on peut retrouver au moins un triplet (a,b,c) correspondant en multipliant par le produit des dénominateurs de x et y (ou leur ppcm).

Or d'après ce qui précède, se donner la paire (x,y) revient à se donner un nombre $t \in \mathbb{Q}$; de plus les conditions $x > 0$ et $y > 0$ (qui s'imposent à nous lorsque l'on prend a, b et c positifs) sont équivalentes à $0 < t < 1$. Nous avons donc une bijection entre les triplets pythagoriciens « à proportionnalité près » et les nombres rationnels entre 0 et 1 ! Il y en a donc bien une infinité.

Pour $t = 1/2$, on trouve $x = 3/5$ et $y = 4/5$ d'où le triplet pythagoricien $(3,4,5)$ que nous connaissions déjà. Pour $t = 1/3$ on tombe sur $(4,3,5)$ qui n'est pas vraiment différent. Par contre pour $t = 1/4$ on obtient $(15,8,17)$ et pour $t = 1/5$ on découvre le triplet $(12,5,13)$.

Ce résultat n'est pas seulement intéressant pour les primitives (voir l'encadré « Triplets pythagoriciens »). Mais pour l'instant, faisons donc le lien avec le calcul des intégrales, dont nous nous sommes écartés temporairement. Rappelons que nous cherchons à calculer

$$\int_p^q \mathrm{F}(\cos(\theta), \sin(\theta))\, \mathrm{d}\theta.$$

La petite étude géométrique ci-dessus nous pousse à introduire le nombre t, qui a l'air de donner des formules simples. Plus précisément, posons

$$t(\theta) = \tan\left(\frac{\theta}{2}\right).$$

Est-ce un bon changement de variables ? Tout d'abord, d'après (**) nous avons

$$\mathrm{F}(\cos(\theta), \sin(\theta)) = \mathrm{F}\left(\frac{1-t(\theta)^2}{1+t(\theta)^2}, \frac{2t(\theta)}{1+t(\theta)^2}\right).$$

Quant à la dérivée $t'(\theta)$, nous sommes chanceux car elle vaut

$$t'(\theta) = \frac{1}{2}\left(1 + \tan\left(\frac{\theta}{2}\right)^2\right) = \frac{1}{2}(1 + t(\theta)^2).$$

On peut donc toujours écrire

$$\begin{aligned}\int_p^q \mathrm{F}(\cos(\theta), \sin(\theta))\, \mathrm{d}\theta &= 2\int_p^q \mathrm{F}\left(\frac{1-t(\theta)^2}{1+t(\theta)^2}, \frac{2t(\theta)}{1+t(\theta)^2}\right) \frac{t'(\theta)}{1+t(\theta)^2}\, \mathrm{d}\theta \\ &= 2\int_{t(p)}^{t(q)} \mathrm{F}\left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\right) \frac{\mathrm{d}t}{1+t^2}.\end{aligned}$$

Lorsque F est une fraction rationnelle trigonométrique, l'expression que nous avons à intégrer est une fraction rationnelle en t. Nous savons donc comment en trouver une primitive.

EXEMPLE 9.17 – Essayons de calculer

$$\mathrm{I} = \int_0^{\frac{\pi}{2}} \frac{\cos(\theta)\, \mathrm{d}\theta}{\cos(\theta) + \sin(\theta)}.$$

Nous avons bien une expression en $\cos(\theta)$ et $\sin(\theta)$. Comme nous l'avons dit, la seule chose à retenir est que poser $t(\theta) = \tan(\frac{\theta}{2})$ est une bonne idée. Écrivons t pour $t(\theta)$, et remplaçons $\cos(\theta)$ par $\frac{1-t^2}{1+t^2}$, puis $\sin(\theta)$ par $\frac{2t}{1+t^2}$. Il vient

$$\frac{\cos(\theta)}{\cos(\theta)+\sin(\theta)} = \frac{1-t^2}{-t^2+2t+1} = \frac{\frac{1}{2}(1-t^2)(1+t^2)}{\frac{1}{2}(-t^2+2t+1)(1+t^2)}.$$

La deuxième manoeuvre est là pour faire apparaître $\frac{1}{2}(1+t^2) = t'(\theta)$. On a donc

$$\mathrm{I} = 2\int_0^1 \frac{(1-t^2)\, \mathrm{d}t}{(-t^2+2t+1)(1+t^2)},$$

le $t'(\theta)\, \mathrm{d}\theta$ devenant $\mathrm{d}t$. Nous avons donc une fraction rationnelle en t.

Le facteur $-t^2+2t+1$ au dénominateur s'annule pour $t = 1 \pm \sqrt{2}$, ce qui permet de factoriser. La décomposition en éléments simples s'écrit alors

$$\frac{1-t^2}{-(t-1-\sqrt{2})(t-1+\sqrt{2})(1+t^2)} = \frac{1}{4(t-1-\sqrt{2})} + \frac{1}{4(t-1+\sqrt{2})} - \frac{t-1}{2(1+t^2)}.$$

D'où la primitive

$$\frac{1}{4}\ln|t-1-\sqrt{2}| + \frac{1}{4}\ln|t-1+\sqrt{2}| - \frac{1}{4}\ln(1+t^2) + \frac{1}{2}\arctan(t).$$

Cette expression vaut 0 pour $t = 0$ (noter que $\ln(\sqrt{2}-1) + \ln(\sqrt{2}+1) = \ln(\sqrt{2}^2 - 1) = \ln(1) = 0$), et elle vaut $\frac{\pi}{8}$ pour $t = 1$. Finalement $\mathrm{I} = \frac{\pi}{4}$.

voir les exercices 2095, 2090

Chapitre 10

Équations différentielles linéaires

Une *équation différentielle* est une équation dans laquelle l'inconnue est une fonction, souvent notée y dans ce chapitre. Par exemple, si f est une fonction quelconque, on peut considérer l'équation différentielle très simple

$$y' = f\,,$$

dont les solutions sont les primitives de f. Nous avons étudié cette équation dans le chapitre sur les intégrales. Autre exemple : l'équation

$$y' = y$$

a pour solutions (définies sur $\mathbb{R}$) les fonctions de la forme $y(x) = ce^x$ avec $c \in \mathbb{R}$, et aucune autre. Ceux qui ont lu l'appendice B ont vu ce résultat à l'occasion du lemme B.4, et nous allons le redémontrer sous peu.

Dans ce chapitre, nous donnons un certain nombre de recettes pour résoudre des équations bien particulières, qui sont parmi celles que l'on rencontre le plus souvent. Nous aurons besoin du calcul de primitives, et aussi de l'algèbre linéaire pour les équations les plus compliquées. L'an prochain vous verrez quelques résultats généraux, notamment sur l'existence et l'unicité des solutions. Cette théorie générale n'est pas nécessaire pour l'instant.

Ce sont les équations de la forme

$$y'(x) = a(x)y(x) + b(x). \qquad \text{(E)}$$

L'équation homogène associée est

$$y'(x) = a(x)y(x). \qquad \text{(H)}$$

En général on a une restriction à $x \in \mathrm{I}$, où $\mathrm{I} \subset \mathbb{R}$ est souvent un intervalle (voir les exemples ci-dessous).
La proposition suivante justifie l'adjectif « linéaire » :

Proposition 10.1 – *L'ensemble $\mathrm{S_H}$ des solutions de l'équation homogène est un espace vectoriel.*
De plus, si y_0 est une solution particulière de l'équation (E), alors n'importe quelle solution y_1 de (E) peut s'écrire

$$y_1 = y_0 + y$$

où y est solution de (H). Réciproquement si y_1 est de cette forme, alors y_1 est solution de (E).

Notons une chose : on peut considérer les solutions y à valeurs dans $\mathbb{R}$, ou bien celles à valeurs dans $\mathbb{C}$. L'espace vectoriel $\mathrm{S_H}$ est lui-même réel ou complexe selon le choix que l'on fait. La théorie est la même dans les deux cas (alors que dans la suite du chapitre, vous serez peut-être surpris de voir que l'on travaille avec $\mathbb{C}$ en général).

Démonstration. Si y et z sont des solutions de (H) alors on calcule tout simplement

$$(y+z)' = y' + z' = ay + az = a(y+z),$$

donc $y + z \in \mathrm{S_H}$. On vérifie également facilement que si $y \in \mathrm{S_H}$ et si λ est un scalaire, alors $\lambda y \in \mathrm{S_H}$. Donc $\mathrm{S_H}$ est un espace vectoriel. (En langage plus savant : l'application $y \mapsto y' - ay$ est linéaire, et $\mathrm{S_H}$ est son noyau.)
Pour la deuxième partie, soient y_0 et y_1 deux solutions de (E), alors

$$(y_1 - y_0)' = y_1' - y_0' = (ay_1 + b) - (ay_0 + b) = a(y_1 - y_0),$$

donc la fonction $y = y_1 - y_0$ est bien solution de (H). Un calcul similaire établit la réciproque. □

Nous connaissons donc la structure de l'ensemble des solutions de (E), et l'on est poussé à commencer par résoudre (H). C'est assez facile.
Voyons d'abord l'idée intuitive (qui va aussi servir de moyen mnémotechnique). On souhaite résoudre l'équation $y' = ay$. On a bien envie d'écrire

$$\frac{y'}{y} = a, \qquad (*)$$

mais alors il faut se restreindre aux fonctions y qui ne s'annulent pas. Ceci étant, on peut intégrer les deux membres de l'équation (*), pour peu que a soit continue. Il vient

$$\int_{x_0}^{x} \frac{y'(t)\,\mathrm{d}t}{y(t)} = [\ln|y(t)|]_{x_0}^{x} = \int_{x_0}^{x} a(t)\,\mathrm{d}t.$$

Ici nous avons supposé que y était définie sur un intervalle I contenant x_0 et x. On va réécrire un peu cette égalité. Posons

$$\alpha(x) = \int_{x_0}^{x} a(t)\,\mathrm{d}t,$$

de sorte que l'on a

$$\ln|y(x)| = \alpha(x) + c_0,$$

où c_0 est une constante. Il vient

$$|y(x)| = e^{\alpha(x)} e^{c_0}.$$

Débarrassons-nous de cette valeur absolue. On a supposé que y ne s'annulait pas, et était définie sur un intervalle ; de plus, on suppose dès le départ que y est dérivable donc continue. D'après le théorème des valeurs intermédiaires, y ne change pas de signe. On a donc ou bien $y(x) = |y(x)|$ pour tout $x \in \mathrm{I}$, ou bien $y(x) = -|y(x)|$ pour tout x. En posant $c = \pm e^{c_0}$, on en conclut que

$$y(x) = ce^{\alpha(x)}, \qquad (**)$$

pour une certaine constante c, et pour tout $x \in \mathrm{I}$.
Nous avons donc montré que les solutions de (H) qui ne s'annulent pas sur un intervalle I sont de la forme (**). On imagine assez difficilement comment une solution pourrait s'annuler, et prendre la forme (**) là où elle ne s'annule pas. C'est en effet impossible, comme le montre le résultat suivant.

Proposition 10.2 – *Supposons que a est continue sur un intervalle I, et que y est une solution définie sur I de l'équation*

$$y'(x) = a(x)y(x) \qquad (x \in \mathrm{I}).$$

Alors il existe une constante c telle que

$$y(x) = ce^{\alpha(x)},$$

où α est une primitive de a, c'est-à-dire que $\alpha' = a$. Réciproquement pour tout c cette expression donne une solution.

Notez que maintenant que nous avons pressenti ce résultat, nous en donnons une démonstration complètement détournée (et très efficace).

Démonstration. Soit α une telle primitive, qui existe puisque a est continue. Considérons la fonction f définie sur I par $f(x) = y(x)e^{-\alpha(x)}$. Alors

$$f'(x) = y'(x)e^{-\alpha(x)} - \alpha'(x)y(x)e^{-\alpha(x)} = (a(x)y(x) - a(x)y(x))e^{-\alpha(x)} = 0.$$

Donc la fonction f est constante sur l'intervalle I, disons $f(x) = c$. D'où le résultat (la réciproque est évidente). □

En particulier, la seule solution qui s'annule est obtenue en prenant $c = 0$, et alors y est la fonction nulle.

Exemple 10.3 – Si on revient à l'équation $y' = y$ sur $\mathrm{I} = \mathbb{R}$, la proposition nous dit que ses solutions sont de la forme $y(x) = ce^x$, comme on le savait.

Exemple 10.4 – Considérons maintenant $y'(x) = xy(x)$, sur $\mathrm{I} = \mathbb{R}$. Pour se rappeler l'énoncé de la proposition, il est très courant de refaire les premières étapes du calcul que nous avons fait en préliminaire. On écrit donc

$$\frac{y'(x)}{y(x)} = x,$$

d'où

$$\ln|y(x)| = \frac{x^2}{2} + \text{constante},$$

puis

$$y(x) = ce^{\frac{x^2}{2}}.$$

(On retient que « la valeur absolue est passée dans c ».)

Nous savons désormais résoudre (H). Pour résoudre (E), reste à trouver *une* solution de (E), et à appliquer la proposition 10.1. À cette fin, il existe un principe général, qui servira dans tout le chapitre (et mérite d'être tenté avec n'importe quelle équation différentielle). Cette méthode porte le nom troublant de *variation des constantes*.

Le principe est le suivant. Après avoir trouvé les solutions de l'équation homogène, on constate qu'elles s'écrivent avec un certain nombre de paramètres (les « constantes »). On peut alors essayer de trouver une solution de l'équation générale en remplaçant ces paramètres par des fonctions (donc en les faisant « varier »).

C'est une recette assez vague, qui s'applique dans de nombreuses situations. Dans le cas des équations linéaires d'ordre 1, les choses sont très simples. Les solutions de (H) sont de la forme $y(x) = ce^{\alpha(x)}$. Pour trouver une solution de (E), on peut alors essayer une fonction de la forme $y(x) = c(x)e^{\alpha(x)}$.

Exemple 10.5 – Cherchons une fonction y telle que

$$y'(x) = y(x) + 1\,. \tag{E}$$

On a vu que les solutions de $y'(x) = y(x)$ sont de la forme $y(x) = ce^x$; d'après le principe de variation de la constante, on a intérêt à chercher une solution sous la forme $y(x) = c(x)e^x$.

On a alors $y'(x) = c'(x)e^x + c(x)e^x = (c'(x) + c(x))e^x$. Si nous remplaçons y' par sa valeur dans (E), on trouve

$$(c'(x) + c(x))e^x = c(x)e^x + 1\,,$$

ce qui revient à $c'(x)e^x = 1$ ou encore $c'(x) = e^{-x}$. On en déduit $c(x) = -e^{-x} +$ constante ; comme on cherche juste une solution (et non pas toutes les solutions), on prend la constante égale à 0. Finalement $c(x) = -e^{-x}$, et $y(x) = -e^{-x}e^x = -1$. La fonction constante égale à -1 est solution de (E), comme on aurait pu le remarquer tout de suite !

Pour finir le travail, appliquons la proposition 10.1. Elle affirme que la « solution générale » de (E) (selon l'expression consacrée) est de la forme $-1 + ce^x$, avec $c \in \mathbb{R}$.

Dans le cas qui nous préoccupe des équations linéaires d'ordre 1, on a en fait le résultat suivant.

Lemme 10.6 – *Pour les équations linéaires d'ordre* 1*, la méthode de la variation de la constante fonctionne toujours, et se ramène à un calcul de primitive.*

C'est pourquoi vous entendrez les gens parler « d'intégrer » une équation différentielle, au lieu de la « résoudre ».

Démonstration. Les solutions de $y' = ay$, lorsque a est continue, sont de la forme $y(x) = ce^{\alpha(x)}$, où $\alpha'(x) = a(x)$. Cherchons une solution de l'équation $y' = ay + b$ sous la forme $y(x) = c(x)e^{\alpha(x)}$.

On a $y'(x) = (c'(x) + a(x)c(x))e^{\alpha(x)} = a(x)y(x) + c'(x)e^{\alpha(x)}$. Donc l'équation $y' = ay + b$ revient à $c'(x)e^{\alpha(x)} = b(x)$, ou encore $c'(x) = b(x)e^{-\alpha(x)}$. On s'est donc bien ramené à calculer une primitive de $b(x)e^{-\alpha(x)}$. □

Exemple 10.7 – Prenons l'équation

$$y'(x) = \frac{y(x)}{1+x^2} + e^{\arctan(x)}\,. \tag{E}$$

Commençons par l'équation homogène

$$y'(x) = \frac{y(x)}{1+x^2} \tag{H}$$

qui donne

$$\frac{y'(x)}{y(x)} = \frac{1}{1+x^2}\,,$$

d'où

$$\ln|y(x)| = \arctan(x) + c\,,$$

et

$$y(x) = ce^{\arctan(x)}\,.$$

Maintenant, cherchons une solution particulière de (E), sous la forme $y(x) = c(x)e^{\arctan(x)}$. On peut faire un calcul direct de y' et remplacer dans (E) ; on peut aussi si l'on préfère retenir la formule obtenue dans la démonstration du lemme ; enfin on peut aussi redémontrer rapidement le lemme, lorsque le cas général est plus clair que le cas particulier considéré (dériver $c(x)e^{\arctan(x)}$ n'est pas tellement plus agréable que dériver $c(x)e^{\alpha(x)}$). Bref, on constate que $y(x) = c(x)e^{\arctan(x)}$ est solution de (E) lorsque

$$c'(x) = e^{\arctan(x)}e^{-\arctan(x)} = 1\,.$$

On prend donc $c(x) = x$, et alors $y(x) = xe^{\arctan(x)}$. On peut vérifier rapidement que c'est bien une solution de (E).

La solution générale de (E) est alors

$$y(x) = xe^{\arctan(x)} + ce^{\arctan(x)}\,,$$

où c est une constante.

voir l'exercice 4054

Sans se restreindre à l'ordre 1, une *équation différentielle linéaire* est par définition de la forme

$$y^{(n+1)}(x) = a_n(x)y^{(n)}(x) + \cdots + a_0(x)y(x) + b(x). \qquad \text{(E)}$$

L'équation homogène associée, sans surprise, est

$$y^{(n+1)}(x) = a_n(x)y^{(n)}(x) + \cdots + a_0(x)y(x). \qquad \text{(H)}$$

La structure générale des solutions ne change pas :

Lemme 10.8 – *La proposition 10.1 est valable pour les équations d'ordre supérieur.*

Démonstration. Si y est une fonction dérivable $(n+1)$ fois sur un ensemble I, alors on note $D(y)$ la fonction

$$D(y) = y^{(n+1)} - a_n y^{(n)} - \cdots - a_0 y\,;$$

alors $D(y)$ est encore une fonction définie sur I. On vérifie tout de suite que $D(y+z) = D(y) + D(z)$ et $D(\lambda y) = \lambda D(y)$ si λ est une constante. Comme S_H est l'ensemble des fonctions y telles que $D(y) = 0$, on voit de suite que c'est un espace vectoriel.

En d'autres termes, on peut voir D comme une application $E \to F$ où E est l'espace vectoriel des fonctions dérivables $(n+1)$ fois sur I, et F est l'espace vectoriel de toutes les fonctions sur I ; alors $S_H = \ker(D)$, c'est donc un espace vectoriel.

Le reste de la proposition 10.1 se démontre dans ce cas plus général *mutatis mutandis*. □

Nous avons vu comment résoudre les équations d'ordre 1 dans le début du chapitre. Il y a un autre cas particulier que l'on sait traiter : celui des équations « à coefficients constants », c'est-à-dire de la forme

$$y^{(n+1)}(x) = a_n y^{(n)}(x) + \cdots + a_0 y(x) + b(x), \qquad \text{(E)}$$

l'équation homogène associée étant

$$y^{(n+1)}(x) = a_n y^{(n)}(x) + \cdots + a_0 y(x). \qquad \text{(H)}$$

On commence par une remarque simple : il est facile de trouver des solutions de la forme $y(x) = e^{\lambda x}$. En effet dans ce cas on a $y'(x) = \lambda e^{\lambda x}$, puis $y''(x) = \lambda^2 e^{\lambda x}$, et on voit immédiatement que $y^{(k)}(x) = \lambda^k e^{\lambda x}$ pour tout $k \geq 0$. Si nous insérons ceci dans l'équation (H), on obtient

$$\lambda^{n+1} e^{\lambda x} = [a_n \lambda^n + \cdots + a_1 \lambda + a_0]\, e^{\lambda x},$$

ce qui revient en simplifiant l'exponentielle à $\chi(\lambda) = 0$, en posant

$$\chi(X) = X^{n+1} - [a_n X^n + \cdots + a_1 X + a_0].$$

On appelle $\chi(X)$ le *polynôme caractéristique* de l'équation (H).

Puisque S_H est un espace vectoriel, on peut obtenir d'autres solutions en considérant des combinaisons linéaires, c'est-à-dire des fonctions de la forme

$$y(x) = c_1 e^{\lambda_1 x} + \cdots + c_m e^{\lambda_m x},$$

où $\lambda_1, \ldots, \lambda_m$ sont des racines du polynôme caractéristique. Nous allons maintenant montrer que, sous l'hypothèse que ce polynôme admet $n+1$ racines distinctes, il n'y a pas d'autres solutions :

Proposition 10.9 – *Supposons que le polynôme caractéristique χ admette $n+1$ racines distinctes $\lambda_0, \lambda_1, \ldots, \lambda_n$ dans $\mathbb{C}$. Alors toute solution y à valeurs complexes de l'équation homogène, définie sur un intervalle, est de la forme*

$$y(x) = c_0 e^{\lambda_0 x} + \cdots + c_n e^{\lambda_n x},$$

où $c_k \in \mathbb{C}$. En particulier une telle solution peut s'étendre en une fonction définie sur $\mathbb{R}$ tout entier.

Avant de donner la démonstration, voyons tout de suite quelques exemples simples.

Exemple 10.10 – Considérons l'équation homogène

$$y'' = y.$$

Le polynôme caractéristique est $\chi(X) = X^2 - 1 = (X-1)(X+1)$. La solution générale est donc

$$y(x) = c_1 e^x + c_2 e^{-x}.$$

Exemple 10.11 – Voyons maintenant

$$y'' = -y.$$

Cette fois le polynôme caractéristique est $\chi(X) = X^2 + 1 = (X+i)(X-i)$. La solution générale est donc

$$y(x) = c_1 e^{ix} + c_2 e^{-ix}. \qquad (*)$$

Mais cette fois-ci, on peut se demander quelle forme particulière est prise par les solutions à valeurs dans $\mathbb{R}$. Écrivons simplement que $y(x) \in \mathbb{R}$ si et seulement si $y(x) = \mathfrak{Re}(y(x))$. En prenant la partie réelle de (*), on constate alors que $y(x)$ est de la forme

$$y(x) = a\cos(x) + b\sin(x), \qquad (**)$$

où a et b sont des constantes réelles. (Si vous faites le calcul vous verrez que l'on a précisément

$$a = \mathfrak{Re}(c_1) + \mathfrak{Re}(c_2) \quad \text{et} \quad b = \mathrm{Im}(c_2) - \mathrm{Im}(c_1),$$

mais ces valeurs importent peu.) Réciproquement, toute fonction de la forme (**) est solution de (H), clairement. Finalement l'expression (**) est la forme générale des solutions de l'équation $y'' = -y$ qui sont à valeurs dans $\mathbb{R}$.

On peut retenir que les solutions réelles s'obtiennent en prenant les parties réelle et imaginaire des solutions complexes.

Passons à la démonstration de la proposition.

Démonstration. Soit y une fonction définie sur un intervalle de $\mathbb{R}$, à valeurs dans $\mathbb{C}$. Définissons alors

$$Y(x) = \begin{pmatrix} y(x) \\ y'(x) \\ \vdots \\ y^{(n)}(x) \end{pmatrix}, \qquad (*)$$

de sorte que

$$Y'(x) = \begin{pmatrix} y'(x) \\ y''(x) \\ \vdots \\ y^{(n+1)}(x) \end{pmatrix}.$$

Dans ces conditions, la fonction y est solution de (H) exactement lorsque

$$Y'(x) = \begin{pmatrix} y'(x) \\ y''(x) \\ \vdots \\ a_n y^{(n)}(x) + a_{n-1} y^{(n-1)}(x) + \cdots + a_0 y(x) \end{pmatrix}.$$

Il se trouve que l'on peut écrire ceci à l'aide d'une matrice. Posons

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_0 & a_1 & a_2 & \cdots & a_n \end{pmatrix}.$$

Alors y est solution de (H) si et seulement si Y est solution de

$$Y'(x) = AY(x). \qquad \text{(V)}$$

Notons d'ailleurs que si une fonction Y quelconque est solution de (V), alors elle doit être de la forme (*).

Conclusion : avec cette astuce (l'introduction de la bonne matrice A), nous avons ramené l'étude de l'équation (H) à l'étude de l'équation (V), qui a l'avantage d'être d'ordre 1, même si les fonctions en jeu sont cette fois à valeurs vectorielles.

L'idée est de diagonaliser A (dans la suite du chapitre nous verrons que c'est systématiquement la chose à faire). Si χ_A est le polynôme caractéristique de A, et si χ est le polynôme caractéristique de l'équation différentielle, il se trouve que l'on a

$$\chi_A = (-1)^{n+1}\chi.$$

Vous montrerez ceci, à titre d'exercice. En particulier, χ_A et χ ont les mêmes racines, et par hypothèse ce sont $\lambda_0, \lambda_1, \ldots, \lambda_n$, qui sont distinctes. D'après le corollaire 16.20, la matrice A est diagonalisable.

Nous savons donc qu'il existe une matrice P telle que

$$P^{-1}AP = \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Appelons D cette matrice diagonale, de sorte que $A = PDP^{-1}$. L'équation différentielle peut alors s'écrire

$$\begin{aligned} Y'(x) = PDP^{-1}Y(x) &\Longleftrightarrow P^{-1}Y'(x) = DP^{-1}Y(x) \\ &\Longleftrightarrow Z'(x) = DZ(x), \end{aligned}$$

en posant $Z(x) = P^{-1}Y(x)$.

Comme D est diagonale, cette nouvelle équation est très simple. Écrivons

$$Z(x) = \begin{pmatrix} z_0(x) \\ \vdots \\ z_n(x) \end{pmatrix}.$$

(Si on calculait P, on pourrait exprimer z_k en fonction de y, mais nous n'aurons même pas besoin de faire ce calcul.) L'équation $Z'(x) = DZ(x)$ s'écrit $z_k'(x) = \lambda_k z_k(x)$ pour chaque k, équation que l'on sait résoudre : on a $z_k(x) = c_k e^{\lambda_k x}$.

Pour récupérer Y, et donc y, on utilise $Y = PZ$. On constate bien que $y(x)$ est une combinaison linéaire des $e^{\lambda_k x}$, comme annoncé. □

Pour résoudre complètement l'équation (E), il faut savoir trouver une solution particulière. À cette fin, on peut appliquer des techniques du type « variation des constantes », mais les calculs sont souvent compliqués. Il convient de connaître un certain nombre d'astuces, et elles seront explorées dans les exercices. *voir l'exercice 4055, sauf le (2)*

– Fin de la première lecture –

Un système d'équations différentielles linéaires, à coefficients constants, est par définition une équation différentielle de la forme

$$Y'(x) = AY(x) + B(x), \qquad \text{(E)}$$

où Y est une fonction à valeurs dans $\mathbb{C}^n$, ainsi que B, et A est une matrice $n \times n$. L'équation homogène associée est

$$Y'(x) = AY(x). \qquad \text{(H)}$$

Exemple 10.12 – Si l'on souhaite trouver deux fonctions y_1 et y_2 telles que

$$\begin{cases} y_1'(x) = -15y_1(x) + 44y_2(x) \\ y_2'(x) = -10y_1(x) + 27y_2(x), \end{cases}$$

alors il s'agit bien d'un système d'équations différentielles linéaires. En effet en notant

$$Y(x) = \begin{pmatrix} y_1(x) \\ y_2(x) \end{pmatrix} \quad \text{et} \quad A = \begin{pmatrix} -15 & 44 \\ -10 & 27 \end{pmatrix},$$

alors on cherche bien à résoudre $Y'(x) = AY(x)$.

Noter qu'en pratique les notations peuvent être très différentes, par exemple le système peut se présenter sous la forme (parfaitement équivalente)

$$\begin{cases} x'(t) = -15x(t) + 44y(t) \\ y'(t) = -10x(t) + 27y(t). \end{cases}$$

Ça sera notamment le cas si l'on pense à t comme au « temps », et à $t \mapsto (x(t), y(t))$ comme à une courbe. Toutefois dans ce chapitre nous garderons des notations uniformes.

Exemple 10.13 – Au cours de la démonstration de la proposition 10.9, nous avons vu que les équations linéaires d'ordre supérieur, à coefficients constants, peuvent se ramener à un système (très particulier). Tout ce que nous allons maintenant démontrer sur les systèmes s'applique donc à cette situation.

De nouveau, la proposition 10.1 s'applique : le soin vous est laissé de faire cette démonstration très facile. Nous allons surtout nous intéresser aux équations homogènes, dans un premier temps du moins, et voir quelques techniques pour l'équation (E) dans les exercices.

La méthode pour résoudre l'équation homogène est simple et tient en un mot : diagonaliser. Plus précisément :

1. On commence par tenter de diagonaliser A, donc de trouver P telle que la matrice $D = P^{-1}AP$ est diagonale. Si c'est impossible, on cherchera une matrice P telle que $P^{-1}AP$ est la plus simple possible (en première année on vous donnera des indications pour ça).
2. L'équation $Y'(x) = AY(x)$, puisque $A = PDP^{-1}$, se réécrit de la manière suivante :
$$\begin{aligned} Y'(x) = PDP^{-1}Y(x) &\iff P^{-1}Y'(x) = DP^{-1}Y(x) \\ &\iff Z'(x) = DZ(x), \end{aligned}$$
en posant $Z(x) = P^{-1}Y(x)$.
3. On note
$$Z(x) = \begin{pmatrix} z_1(x) \\ \vdots \\ z_n(x) \end{pmatrix},$$
puis on résout l'équation $Z'(x) = DZ(x)$. Lorsque D est diagonale, ceci revient à résoudre une équation simple pour chaque z_k, et on trouve z_k immédiatement.
4. On retrouve $Y(x)$ par la formule $Y(x) = PZ(x)$.

Il est important de noter que *l'on n'a pas besoin de calculer l'inverse de la matrice* P, à aucun moment. En effet les techniques de diagonalisation que l'on a vues permettent de trouver P telle que $P^{-1}AP$ est diagonale sans calculer P^{-1}.

Exemple 10.14 – Revenons à l'exemple 10.12. La matrice est

$$A = \begin{pmatrix} -15 & 44 \\ -10 & 27 \end{pmatrix}.$$

Le polynôme caractéristique est $\chi_A = \lambda^2 - 12\lambda + 35 = (\lambda - 7)(\lambda - 5)$, les valeurs propres sont 7 et 5, la matrice est diagonalisable. On cherche les vecteurs propres, on trouve par exemple

$$e_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{et} \quad e_2 = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

associés à 7 et 5 respectivement. On en conclut qu'en posant

$$P = \begin{pmatrix} 2 & 11 \\ 1 & 5 \end{pmatrix},$$

alors

$$P^{-1}AP = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} = D.$$

On pose alors

$$Z(x) = P^{-1}Y(x) = \begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix}.$$

(On n'a pas calculé P^{-1}.) Alors Z satisfait l'équation $Z'(x) = DZ(x)$, ce qui s'écrit

$$\begin{cases} z_1'(x) = 7z_1(x) \\ z_2'(x) = 5z_2(x). \end{cases}$$

On sait bien faire ça : $z_1(x) = c_1e^{7x}$ et $z_2(x) = c_2e^{5x}$.

Pour finir $Y(x) = PZ(x)$ donc

$$\begin{cases} y_1(x) = 2c_1e^{7x} + 11c_2e^{5x} \\ y_2(x) = c_1e^{7x} + 5c_2e^{5x} \end{cases}.$$

Exemple 10.15 – Parfois la matrice A n'est pas diagonalisable (ça sera évidemment plus rare). Par exemple on peut considérer le système

$$\begin{cases} y_1'(x) = y_1(x) + y_2(x) \\ y_2'(x) = -y_1(x) + 3y_2(x). \end{cases}$$

Le polynôme caractéristique est $(\lambda - 2)^2$, et l'espace propre associé à l'unique valeur propre 2 est de dimension 1, avec pour base par exemple

$$e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

La matrice n'est pas diagonalisable. Prenons n'importe quel vecteur e_2 tel que e_1, e_2 est une base, par exemple

$$e_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

et soit P la matrice dont les colonnes sont e_1 et e_2. Calculons

$$P^{-1}AP = \begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix} = D.$$

(Un peu de réflexion montre que, quel que soit notre choix pour e_2, la matrice $P^{-1}AP$ doit être triangulaire, et en calculant la trace ou le déterminant on s'assure que la diagonale doit être $(2, 2)$. La méthode qui suit ne dépend pas vraiment du choix de e_2.)

On continue d'appliquer la méthode. On pose

$$Z(x) = P^{-1}Y(x) = \begin{pmatrix} z_1(x) \\ z_2(x) \end{pmatrix}.$$

On a $Z'(x) = DZ(x)$, ce qui s'écrit

$$\begin{cases} z_1'(x) = 2z_1(x) - z_2(x) \\ z_2'(x) = 2z_2(x) \end{cases}.$$

On peut résoudre la deuxième d'abord : $z_2(x) = c_2e^{2x}$. En remplaçant dans la première, nous obtenons

$$z_1'(x) = 2z_1(x) - c_2e^{2x}. \qquad (*)$$

C'est une équation d'ordre 1, que l'on sait résoudre ! L'équation homogène est $z_1'(x) = 2z_1(x)$ qui a pour solutions les fonctions de la forme c_1e^{2x}. En appliquant la méthode de variation de la constante, on cherche une solution de (*) de la forme $c(x)e^{2x}$; on constate que l'équation revient à $c'(x) = -c_2e^{2x}e^{-2x} = -c_2$. On prend $c(x) = -c_2x$ ce qui donne la solution $-c_2xe^{2x}$. Finalement la solution générique de (*) est

$$z_1(x) = c_1e^{2x} - c_2xe^{2x}.$$

On retrouve finalement y_1 et y_2 d'après la relation $Y(x) = PZ(x)$, c'est-à-dire $y_1(x) = z_1(x) + z_2(x) = c_1e^{2x} + c_2(1-x)e^{2x}$ et $y_2(x) = z_1(x) = c_1e^{2x} - c_2xe^{2x}$.

voir l'exercice 4064

Nous avons toutes les clefs en main pour résoudre les systèmes d'équations différentielles en pratique. À présent, nous allons montrer quelques propriétés qualitatives : le calcul de la dimension de l'espace vectoriel des solutions, leur forme générale, et les intervalles sur lesquels on peut les définir. En deux mots, nous trouverons que pour un système $n \times n$, les solutions forment un espace de dimension n, on peut les exprimer avec des exponentielles et des polynômes, et par conséquent on peut naturellement les étendre à $\mathbb{R}$ tout entier. Les démonstrations font surtout appel à de l'algèbre linéaire.

Commençons par un petit lemme de calcul.

Lemme 10.16 – *Soit $f(x) = P(x)e^{\lambda x}$, où P est un polynôme. Alors f possède une primitive de la forme $Q(x)e^{\lambda x}$, où Q est encore un polynôme. Si $\lambda \neq 0$, alors $\deg Q \leq \deg P$; si $\lambda = 0$, alors $\deg Q = \deg P + 1$.*

Démonstration. Le cas $\lambda = 0$ est évident donc on se tourne vers $\lambda \neq 0$. La dérivée de $Q(x)e^{\lambda x}$ est $(\lambda Q(x) + Q'(x))e^{\lambda x}$. Il suffit donc de montrer que pour tout polynôme P, il existe un polynôme Q de degré $\leq \deg P$ tel que $P = \lambda Q + Q'$.

Pour montrer ceci, soit $n = \deg P$, et soit

$$\begin{array}{rccc} \phi\colon & \mathbb{C}_n[X] & \longrightarrow & \mathbb{C}_n[X] \\ & Q & \mapsto & \lambda Q + Q'. \end{array}$$

C'est une application linéaire ; regardons $\ker(\phi)$. On a $\phi(Q) = 0 \iff \lambda Q = -Q'$, et si $Q \neq 0$ c'est impossible pour des raisons de degré. Donc $\ker(\phi) = \{0\}$. Par suite, ϕ est injective, et donc surjective aussi (corollaire 15.44). Ainsi il existe Q tel que $\phi(Q) = P$. □

Théorème 10.17 – *On considère une équation différentielle de la forme*

$$Y'(x) = AY(x), \tag{H_A}$$

où $A \in M_n(\mathbb{C})$. Alors l'espace vectoriel S_A des solutions est de dimension n.

De plus, les solutions ont la forme suivante. Écrivons

$$Y(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix},$$

et soit

$$\chi_A(\lambda) = (\lambda_1 - \lambda)^{m_1} \cdots (\lambda_s - \lambda)^{m_s}$$

le polynôme caractéristique de A, factorisé sur $\mathbb{C}$. Alors $y_k(x)$ peut s'écrire comme une combinaison linéaire des expressions

$$x^j e^{\lambda_i x}$$

avec $j < m_i$.

En particulier y_k s'étend naturellement en une solution définie sur $\mathbb{R}$.

Démonstration. On va procéder par récurrence sur n. Le cas $n = 1$ est celui des équations linéaires d'ordre 1 que l'on connaît bien, et le théorème est alors évidemment vrai. Supposons donc le théorème démontré pour $n-1$ et montrons-le pour n.

Le point essentiel est de comprendre ce qui se passe lorsqu'on « remplace » la matrice A par une matrice conjuguée $B = P^{-1}AP$. Comme nous l'avons vu dans les exemples, si Y est une solution de $Y' = AY$, alors $Z = P^{-1}Y$ est solution de $Z' = BZ$, et *vice-versa*. De manière plus savante, on a un isomorphisme

$$\begin{array}{ccc} S_A & \longrightarrow & S_B \\ Y & \mapsto & P^{-1}Y. \end{array}$$

L'inverse est donné par $Z \mapsto PZ$, bien sûr. On en déduit que $\dim S_A = \dim S_B$. D'autre part, chaque composante z_k de Z est une combinaison linéaire des composantes y_k de Y (et réciproquement), donc elles ont la même « forme » (combinaisons de certains polynômes et d'exponentielles). Conclusion : il suffit de montrer le théorème pour B, il sera alors vrai pour A.

Choisissons donc P telle que la matrice $B = P^{-1}AP$ est de la forme

$$B = \left(\begin{array}{ccc|c} & & & 0 \\ & A_{n-1} & & \vdots \\ & & & 0 \\ \hline * & \cdots & * & \mu \end{array} \right).$$

C'est possible d'après le corollaire 16.26 (qui dit même que l'on peut prendre A_{n-1} triangulaire inférieure, mais ça ne sera pas utile). Notons que

$$\chi_A(\lambda) = \chi_B(\lambda) = (\mu - \lambda)\chi_{A_{n-1}}(\lambda),$$

comme on le voit en développant par la dernière colonne. En particulier μ est une valeur propre de A, c'est-à-dire que $\mu = \lambda_{i_0}$ pour un certain indice i_0.

Soit donc Z une solution de $Z'(x) = BZ(x)$, et notons

$$Z(x) = \begin{pmatrix} z_1(x) \\ \vdots \\ z_n(x) \end{pmatrix}.$$

Enfin notons

$$Z_{n-1}(x) = \begin{pmatrix} z_1(x) \\ \vdots \\ z_{n-1}(x) \end{pmatrix},$$

c'est-à-dire que Z_{n-1} est obtenu en ne gardant que les $n-1$ premières composantes de Z. D'après la forme de B, on constate que Z_{n-1} vérifie $Z'_{n-1}(x) = A_{n-1}Z_{n-1}(x)$ (c'est ce que l'on voit en ne regardant que les $n-1$ premières équations du système $Z'(x) = BZ(x)$). On a donc une application linéaire

$$\begin{array}{rccc} \phi\colon & S_B & \longrightarrow & S_{A_{n-1}} \\ & Z & \mapsto & Z_{n-1}. \end{array}$$

Nous allons montrer deux choses : d'une part, que ϕ est surjective, et d'autre part que $\dim\ker(\phi) = 1$. En effet, supposons ces deux choses établies, et appliquons le théorème du rang. Nous avons

$$\dim S_B = \dim\ker(\phi) + \dim(\mathrm{Im}(\phi)),$$

et par récurrence, on sait que $\dim \mathrm{Im}(\phi) = \dim S_{A_{n-1}} = n-1$. On a donc $\dim S_B = 1 + (n-1) = n = \dim S_A$. Ceci donne bien le calcul de la dimension au rang n.

Le plus simple est l'étude de $\ker(\phi)$. Si $\phi(Z) = 0$, alors Z est de la forme

$$Z(x) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ z_n \end{pmatrix},$$

et l'équation $Z'(x) = BZ(x)$ équivaut à $z'_n(x) = \mu z_n$; cette dernière équation est un système d'ordre 1, donc ses solutions forment un espace de dimension 1 (et concrètement, $z_n(x) = ce^{\mu x}$). Ceci montre que $\dim\ker(\phi) = 1$.

Montrons que ϕ est surjective. Il faut montrer que, étant données des fonctions $z_1, \ldots, z_{n-1}$ qui forment une solution de $Z'_{n-1}(x) = A_{n-1}Z_{n-1}(x)$, on peut les compléter en une solution Z de $Z'(x) = BZ(x)$ en ajoutant une fonction bien choisie z_n. Or la seule équation du système $Z'(x) = BZ(x)$ faisant intervenir z_n est la dernière, qui est de la forme

$$z'_n(x) = \mu z_n(x) + b(x). \tag{*}$$

On sait étudier les équations d'ordre 1 ; en particulier, une telle équation possède toujours des solutions lorsque b est continue, ce qui est le cas (voir plus bas la forme de b). Donc on peut trouver z_n, et ϕ est surjective. Ceci conclut la démonstration que l'espace des solutions est de dimension n.

Pour finir la démonstration, il reste à établir que les solutions ont la forme annoncée, et par récurrence on le sait pour Z_{n-1} ; reste donc à voir la forme de z_n, ce qui va se faire en regardant de plus près l'équation (*). Puisque les fonctions z_k sont des combinaisons de polynômes et d'exponentielles, pour $1 \leq k \leq n-1$, on a

$$b(x) = \sum_i P_i(x)e^{\lambda_i x}.$$

De plus on a $\deg P_i < m_i$ pour $i \neq i_0$ alors que $\deg P_{i_0} < m_{i_0} - 1$ (rappelez-vous la notation : i_0 est l'indice tel que $\mu = \lambda_{i_0}$, et on utilise le fait que λ_{i_0} comme valeur propre de A_{n-1} a pour multiplicité $m_{i_0} - 1$).

La solution générale de (*) est donc

$$z_n(x) = ce^{\mu x} + c(x)e^{\mu x},$$

où $c(x)$ vérifie

$$c'(x) = b(x)e^{-\mu x} = \sum_i P_i(x)e^{(\lambda_i - \mu)x}.$$

On peut maintenant appliquer le lemme 10.16. Il nous dit que l'on peut choisir $c(x)$ sous la forme

$$c(x) = \sum_i Q_i(x)e^{(\lambda_i - \mu)x},$$

où Q_i est un polynôme, avec $\deg Q_i \leq \deg P_i$, sauf si $i = i_0$, et alors $\deg Q_{i_0} = \deg P_{i_0} + 1$. Notons alors que le produit $c(x)e^{\mu x}$ peut s'écrire

$$c(x)e^{\mu x} = \sum_i Q_i(x)e^{\lambda_i x}.$$

Ceci montre que $z_n(x)$ a précisément la forme annoncée. □

Nous pouvons revisiter les exemples étudiés ci-dessus à la lumière du théorème.

Exemple 10.18 – Le système de l'exemple 10.12 était

$$\begin{cases} y'_1(x) = -15y_1(x) + 44y_2(x) \\ y'_2(x) = -10y_1(x) + 27y_2(x), \end{cases}$$

la solution était donnée dans l'exemple 10.14. On a vu que le polynôme caractéristique était $(\lambda-7)(\lambda-5)$. Le théorème nous affirme alors que y_1 est de la forme

$$y_1(x) = ae^{7x} + be^{5x},$$

où a et b sont des constantes, et que y_2 est de la forme

$$y_2(x) = ce^{7x} + de^{5x}.$$

Attention : le théorème ne dit certainement pas que, réciproquement, on obtient des solutions en prenant a, b, c, d arbitrairement ; d'ailleurs on ne s'attend pas à avoir 4 paramètres libres alors que le théorème nous dit également que l'espace des solutions est de dimension 2.

En fait, nous avions fait les calculs et trouvé que

$$\begin{cases} y_1(x) = 2c_1e^{7x} + 11c_2e^{5x} \\ y_2(x) = c_1e^{7x} + 5c_2e^{5x} \end{cases}$$

où c_1 et c_2 sont des constantes *arbitraires* (il y en a bien 2 !).

Regardons de même l'exemple 10.15, pour lequel le système est

$$\begin{cases} y'_1(x) = y_1(x) + y_2(x) \\ y'_2(x) = -y_1(x) + 3y_2(x). \end{cases}$$

Le polynôme caractéristique est $(\lambda - 2)^2$. Le théorème affirme donc que

$$y_1(x) = ae^{2x} + bxe^{2x},$$

et

$$y_2(x) = ce^{2x} + dxe^{2x},$$

où a, b, c, d sont des constantes ; l'espace des solutions est de dimension 2 donc il doit y avoir des relations entre ces constantes. Nous avions fait les calculs et trouvé que

$$y_1(x) = (c_1 + c_2)e^{2x} - c_2xe^{2x},$$

et que

$$y_2(x) = c_1e^{2x} - c_2xe^{2x},$$

avec deux constantes *arbitraires* c_1 et c_2.

Les résultats prédits par le théorème sont donc cohérents avec ceux que l'on avait obtenus directement. Le théorème à lui seul ne donne pas assez d'information pour résoudre le système, et d'ailleurs on s'en passe. Mais l'intérêt de cet énoncé abstrait est de dire au moins quelque chose dans les cas où un calcul direct n'est pas envisageable, par exemple si la matrice est très grande, où si elle dépend elle-même de paramètres d'une manière compliquée.

Pour se convaincre de l'intérêt du théorème 10.17, voici une conséquence concrète : nous pouvons maintenant résoudre toutes les équations différentielles linéaires d'ordre supérieur, à coefficients constants. Rappelons que dans la proposition 10.9 nous avions seulement traité le cas où les racines du polynôme caractéristique étaient distinctes.

Proposition 10.19 – *On considère l'équation homogène*

$$y^{(n+1)}(x) = a_n y^{(n)}(x) + \cdots + a_0 y(x). \tag{H}$$

Soit

$$\chi(\lambda) = \lambda^{n+1} - [a_n\lambda^n + \cdots + a_0] = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_s)^{m_s}$$

le polynôme caractéristique, factorisé sur $\mathbb{C}$*. Alors toute solution* y *de (H) peut s'écrire de manière unique*

$$y(x) = \sum_{i=1}^{s} \sum_{j=0}^{m_i - 1} c_{ij} x^j e^{\lambda_i x},$$

et réciproquement pour tout choix de coefficients c_{ij}*, une fonction de cette forme est solution.*

Démonstration. On reprend le début de la démonstration de la proposition 10.9 : on pose

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \\ a_0 & a_1 & a_2 & \cdots & a_n \end{pmatrix},$$

et alors on constate que Y est solution de $Y'(x) = AY(x)$ si et seulement si elle est de la forme

$$Y(x) = \begin{pmatrix} y(x) \\ y'(x) \\ \vdots \\ y^{(n)}(x) \end{pmatrix}, \tag{*}$$

avec y solution de (H).

Si on note S_A l'espace des solutions de $Y' = AY$, et S_H l'espace des solutions de (H), alors on a un isomorphisme $S_A \to S_H$ donné par $Y \mapsto y$; l'inverse envoie y sur la fonction Y définie par (*). On en conclut, d'après le théorème 10.17, que $\dim S_H = \dim S_A = n+1$.

Or, d'après le même théorème, on sait que y doit avoir précisément la forme annoncée dans la proposition (rappelons que le polynôme caractéristique de A est, au signe près, le polynôme caractéristique considéré dans l'énoncé).

Il est facile de conclure. Soit e_{ij} la fonction définie sur $\mathbb{R}$ par $e_{ij}(x) = x^j e^{\lambda_i x}$, pour $1 \leq i \leq s$ et $0 \leq j < m_i$. Ces fonctions sont au nombre de $n+1$, qui est le degré du polynôme χ. Considérons alors l'espace vectoriel $E = \mathcal{V}ect(e_{ij})$ (c'est un sous-espace de l'espace de toutes les fonctions $\mathbb{R} \to \mathbb{C}$). Il est donc de dimension $\leq n+1$. Mais on vient de voir que $S_H \subset E$ et que $\dim S_H = n+1$. On en conclut que $S_H = E$ et que $\dim E = n+1$; de plus la famille des e_{ij} doit donc être libre.

On a donc bien montré que chaque $e_{ij} \in S_H$, donc est solution de (H), et que chaque solution de (H) s'écrivait de manière unique comme combinaison linéaire de ces fonctions. □

Exemple 10.20 – Considérons l'équation

$$y^{(3)}(x) + 3y''(x) + 3y'(x) + y(x) = 0.$$

Le polynôme caractéristique est $\lambda^3 + 3\lambda^2 + 3\lambda + 1 = (\lambda+1)^3$. On en conclut que les solutions sont précisément les fonctions de la forme

$$(a + bx + cx^2)e^{-x},$$

où a, b et c sont des constantes arbitraires.

reprendre le (2) de l'exercice 4055

Le lecteur ayant parcouru la deuxième partie de l'appendice « L'exponentielle » connait les exponentielles de matrices. Rappelons que si A est une matrice, alors par définition

$$e^{A} = \sum_{k=0}^{+\infty} \frac{1}{k!} A^{k}.$$

La proposition B.24 affirme alors que la fonction $\gamma(x) = e^{xA}$ vérifie $\gamma'(x) = Ae^{xA}$. On aperçoit alors une nouvelle façon de produire des solutions de nos systèmes d'équations différentielles : en effet, pour tout vecteur $v \in \mathbb{C}^n$, si on pose $Y(x) = e^{xA}v$, alors $Y'(x) = Ae^{xA}v = AY(x)$.

Et toutes les solutions sont en fait de cette forme :

Proposition 10.21 – *Soit* Y *une solution du système* $Y'(x) = AY(x)$*. Alors* Y *est de la forme* $Y(x) = e^{xA}v$ *pour un certain vecteur* v *; par suite* $v = Y(0)$*. Réciproquement toutes les fonctions de cette forme sont solutions.*

En particulier, une solution Y *de ce système est entièrement déterminée par le vecteur* $Y(0)$.

Démonstration. Nous avons déjà vu la partie « réciproque ». On doit montrer que toute solution Y est de cette forme.

Il y a deux démonstrations faciles. La première consiste à adapter la démonstration de la proposition 10.2, en remplaçant l'exponentielle usuelle par l'exponentielle de matrice. Il n'y a presque rien à changer (et le peu qu'il y a à changer a en fait été vu lors de la démonstration de la proposition B.24).

Voici la deuxième démonstration. On considère l'application $\phi\colon S_A \to \mathbb{C}^n$ définie par $\phi(Y) = Y(0)$ (comme d'habitude S_A est l'espace vectoriel des solutions du système). L'application ϕ est évidemment linéaire. De plus, elle est surjective, puisque pour tout $v \in \mathbb{C}^n$, la solution $Y \in S_A$ définie par $Y(x) = e^{xA}v$ vérifie $Y(0) = v$. D'après le théorème 10.17, la dimension de S_A est n, et on en conclut que ϕ est injective également. Ceci montre que Y est déterminée par $v = Y(0)$. Ainsi, la *seule* solution Y telle que $Y(0) = v$ est $Y(x) = e^{xA}v$. □

Il semblerait que nous ayons donné une formule simple pour exprimer les solutions de $Y'(x) = AY(x)$, et en un sens c'est le cas. Mais il ne faudrait pas croire que cette formule va nous dispenser de la méthode que nous connaissons pour résoudre un tel système en pratique. En effet, le calcul de e^A est compliqué, et se fait... en diagonalisant A (voir l'exemple B.23). On ne gagne pas vraiment de temps en procédant ainsi.

Par contre, le fait qu'une solution Y est déterminée par Y(0) est nouveau. Vous verrez l'an prochain qu'il y a là un phénomène général dans la théorie des équations différentielles.

Exemple 10.22 – De nouveau, retournons à l'exemple 10.12, donc au cas où

$$A = \begin{pmatrix} -15 & 44 \\ -10 & 27 \end{pmatrix}.$$

Les solutions de $Y'(x) = AY(x)$ sont, d'après la proposition, de la forme $e^{xA}v$, et nous devons donc commencer par calculer e^{xA}. On a vu (exemple 10.14) que

$$P^{-1}AP = \begin{pmatrix} 7 & 0 \\ 0 & 5 \end{pmatrix} = D \quad \text{avec} \quad P = \begin{pmatrix} 2 & 11 \\ 1 & 5 \end{pmatrix}.$$

On en tire $xA = P(xD)P^{-1}$ et

$$e^{xA} = Pe^{xD}P^{-1},$$

voir la proposition B.22. De plus comme xD est diagonale, on a évidemment

$$e^{xD} = \begin{pmatrix} e^{7x} & 0 \\ 0 & e^{5x} \end{pmatrix}.$$

On veut s'éviter le calcul de P^{-1} (rappelons que nous n'avons pas eu besoin de faire ce calcul pour diagonaliser A). On va donc garder le résultat sous cette forme, et écrire qu'il existe un vecteur v tel que

$$Y(x) = e^{xA}v = Pe^{xD}P^{-1}v = \begin{pmatrix} 2e^{7x} & 11e^{5x} \\ e^{7x} & 5e^{5x} \end{pmatrix} w,$$

en posant $w = P^{-1}v$. On sait que v peut être choisi librement, donc w peut être choisi librement ; en posant

$$w = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

on retrouve le résultat de l'exemple 10.14, à savoir que Y est de la forme

$$Y(x) = \begin{pmatrix} 2c_1e^{7x} + 11c_2e^{5x} \\ c_1e^{7x} + 5c_2e^{5x} \end{pmatrix}.$$

Cette méthode n'est ni plus rapide, ni plus lente que la précédente.

reprendre les systèmes précédents par cette méthode

Troisième partie

Algèbre

Chapitre 11

Polynômes

Dans ce chapitre, et dans toute la partie « Algèbre », la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$. Le lecteur ayant assimilé la définition 2.17 peut prendre pour $\mathbb{K}$ n'importe quel corps.

DÉFINITION 11.1 – Donnons-nous un symbole X. Un *polynôme en* X *à coefficients dans* $\mathbb{K}$ est une expression formelle

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$$

avec $a_n \neq 0$. L'entier n est appelé le degré du polynôme.

L'ensemble de ces polynômes est noté $\mathbb{K}[X]$, et le sous-ensemble des polynômes de degré $\leq n$ est noté $\mathbb{K}_n[X]$.

Les termes « symbole » et « expression formelle » sont à comprendre de manière intuitive : disons qu'un polynôme est une écriture. Si vous trouvez ça insatisfaisant, essayez l'encadré « Définition complète des polynômes ». Pour tout le monde, précisons que l'on autorise aussi le « polynôme nul », dont tous les coefficients sont 0 (et pour lequel on prend parfois la convention de dire que son degré est $-\infty$ afin que certaines formules restent vraies).

Définition complète des polynômes

À un interlocuteur exigeant et rigoureux, ou à un ordinateur, nous dirions qu'un polynôme est défini par ses coefficients, et puis nous indiquerions les règles de calcul sur ces coefficients. Voici les détails.

Considérons les fonctions $\mathbb{N} \to \mathbb{K}$, autrement dit les suites d'éléments de $\mathbb{K}$. Une telle suite a sera notée

$$a = (a_0, a_1, a_2, \ldots),$$

avec $a_n = a(n)$.

On définit une addition $\oplus$ le plus simplement du monde : si $b = (b_0, b_1, \ldots)$, alors nous définissons

$$a \oplus b = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots).$$

Définissons maintenant une multiplication $\otimes$, qui paraît bien plus tordue : $a \otimes b = (c_0, c_1, c_2, \ldots)$, où

$$c_n = \sum_{p=0}^{n} a_p b_{n-p}.$$

(On appelle ceci la « formule de Cartan ».)

On peut vérifier directement que l'ensemble des fonctions $\mathbb{N} \to \mathbb{K}$, avec ces opérations, est un anneau commutatif (cf. définition 2.17).

Première remarque : en identifiant $x \in \mathbb{K}$ avec la suite $(x, 0, 0, 0, \ldots)$, on peut considérer que $\mathbb{K}$ est contenu dans cet ensemble de suites. De plus cette identification est compatible avec l'addition et la multiplication sur $\mathbb{K}$.

Soit maintenant $X = (0, 1, 0, 0, 0, \ldots)$, c'est-à-dire $X(n) = 0$ sauf si $n = 1$, et $X(1) = 1$. Essayons quelques calculs :

$$X^2 = X \otimes X = (0, 0, 1, 0, 0, \ldots),$$

$$X^3 = X \otimes X^2 = (0, 0, 0, 1, 0, \ldots),$$

et de même on constate que X^n est représenté par une suite de 0, sauf à la position n où l'on trouve un 1 (faire une récurrence simple).

Finalement, soient $a_0, a_1, \ldots, a_n$ des éléments de $\mathbb{K}$. Si l'on calcule $a_0 \oplus a_1 \otimes X \oplus \cdots \oplus a_n \otimes X^n$, on trouve

$$(a_0, a_1, \ldots, a_n, 0, 0, 0, \ldots).$$

On peut finalement définir $\mathbb{K}[X]$ comme étant l'ensemble des suites $a : \mathbb{N} \to \mathbb{K}$ telles que $a(k) = 0$ pour tous les k supérieurs à un certain $n \in \mathbb{N}$ appelé le degré. Les opérations sont celles ci-dessus, et on va écrire $P+Q$ et PQ au lieu de $P \oplus Q$ et $P \otimes Q$, pour simplifier. On *vérifie* alors que tout polynôme P s'écrit de manière unique

$$P = a_0 + a_1 X + \cdots + a_n X^n.$$

Par analogie, une suite *quelconque* $a : \mathbb{N} \to \mathbb{K}$ peut être notée

$$\sum_{n \geq 0} a_n X^n,$$

et appelée une *série formelle*, lorsqu'on veut faire référence aux opérations d'addition et de multiplication que l'on vient de définir. Attention cependant : cette notation ne doit pas donner l'illusion d'une somme « infinie » ou d'un passage à la limite.

On note $\mathbb{K}[[X]]$ l'ensemble des séries formelles.

Par exemple, $P = 1 + 5X + 3X^2$ est un polynôme de $\mathbb{Q}[X]$, et $Q = 7 + X + iX^2 + (2 - 4i)X^3 \in \mathbb{C}[X]$. Noter que $3 - X^2$ est aussi un polynôme, en notation raccourcie pour $3 + 0X + (-1)X^2$.

Lorsque l'on dispose d'un polynôme $P \in \mathbb{K}[X]$ et d'un élement $x \in \mathbb{K}$, on peut donner un sens à $P(x)$. Sans surprise, si

$$P = a_0 + a_1 X + \cdots + a_n X^n,$$

alors

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

On dit que l'on évalue P en x. Si $P = 1 + X^2$, alors $P(-2) = 1 + (-2)^2 = 5$, par exemple. Cette opération est tellement commune que l'on note souvent $P(X)$ (au lieu de P tout simplement) pour un élément de $\mathbb{K}[X]$, afin de rappeler cette possibilité d'évaluer.

Un polynôme donne naissance à plusieurs fonctions. Prenons $P = -12X^3 + 7X^5$; on peut considérer la fonction

$$\begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \mapsto & P(x) = -12x^3 + 7x^5. \end{array}$$

Mais on peut aussi regarder

$$\begin{array}{ccc} \mathbb{C} & \longrightarrow & \mathbb{C} \\ z & \mapsto & P(z) = -12z^3 + 7z^5, \end{array}$$

et il y aurait aussi la fonction $[0,1] \to \mathbb{R}$ qui à x associe $P(x)$, etc.

On peut additionner les polynômes de façon naturelle, coefficient par coefficient. Par exemple si $P = 1 + X^2$ et $Q = -1 - 3X + X^2$, alors $P + Q = -3X + 2X^2 = Q + P$.

La règle de commutativité $P+Q = Q+P$ est générale, et permet d'écrire les sommes dans n'importe quel ordre. En prenant $P = 4X^3$, $Q = 5X$ et $R = 2$, on a ainsi $P + Q + R = 4X^3 + 5X + 2 = 2 + 5X + 4X^3 = R + Q + P$. Maintenant que cette observation est faite, nous pourrons écrire les termes des polynômes dans l'ordre qui nous convient le mieux.

On a également une multiplication : sans surprise, on développe puis on regroupe les termes. Si $P = X^2 - X + 2$ et $Q = 2X^5 - 3$ alors

$$P \cdot Q = 2X^7 - 2X^6 + 4X^5 - 3X^2 + 3X - 6 = Q \cdot P.$$

De nouveau, la règle $P \cdot Q = Q \cdot P$ est générale. (Là encore, les amateurs de détails se rapporteront à l'encadré précédent.)

Après quelques essais, on se rend compte qu'on ne peut pas toujours diviser : par exemple si $P = X^2 - 1$ et $Q = X + 2$, il n'y a pas de polynôme R qui mériterait de s'appeler $\frac{P}{Q}$, c'est-à-dire qu'il n'y a pas de polynôme R tel que $QR = P$. Pour montrer ceci, on note que si R existait, il serait de degré 1, disons $R = aX + b$. Or l'équation $QR = P$ donne en développant :

$$aX^2 + (2a + b)X + 2b = X^2 - 1,$$

et en comparant les coefficients, on obtient $a = 1$, $b = \frac{-1}{2}$, et $2a + b = \frac{3}{2} = 0$, contradiction.

Parfois, on peut avoir de la chance : pour le même P, et pour $Q = X + 1$, on a $P = X^2 - 1 = (X - 1)(X + 1) = (X - 1)Q$ donc $\frac{P}{Q} = X - 1$.

En général on dit que Q *divise* P lorsqu'il existe un polynôme R tel que $P = QR$. Dans ce cas, et dans ce cas seulement, on pourra noter $R = \frac{P}{Q}$ (ce qui a un sens car R est unique : exercice !). On utilise la notation $Q \mid P$ pour indiquer que Q divise P. Il faut se méfier de cette notation (standard malheureusement) qui apparaît symétrique alors que les rôles de P et Q sont très différents.

La situation des polynômes est très similaire à celle des nombres entiers : on peut parfois diviser un entier par un autre, parfois « ça ne tombe pas juste ». Dans tout ce chapitre on va insister sur les similarités, et nous commençons par les divisions euclidiennes.

Rappelons que, si a et $b \neq 0$ sont des nombres entiers, il existe deux nombres entiers q (le quotient) et r (le reste), uniques, tels que

$$a = bq + r,$$

avec $0 \leq r < b$.

Proposition 11.2 – *Soient* A *et* $B \neq 0$ *deux polynômes de* $\mathbb{K}[X]$. *Alors il existe deux polynômes* Q *(le quotient) et* R *(le reste), uniques, tels que*

$$A = BQ + R,$$

avec $\deg R < \deg B$.

Démonstration. Montrons l'unicité. Si $A = BQ + R$ et $A = BQ' + R'$, en faisant la différence on obtient

$$B(Q - Q') = R' - R.$$

Le degré de $R' - R$ est $< \deg B$. On en déduit que $Q - Q' = 0$, sinon le degré de $B(Q - Q')$ serait $\geq \deg B$. Donc $Q = Q'$ et par suite $R = R'$.

Pour l'existence de Q et R, nous allons donner directement une méthode de calcul. □

Exemple 11.3 – Prenons $A = 4X^3 - 2X^2 + 1$ et $B = X^2 + X + 1$. On commence par présenter la division comme pour les nombres entiers :

$$\begin{array}{cccc|l} 4X^3 & -2X^2 & & +1 & X^2+X+1 \\ \hline \end{array}$$

Puis on évalue « en $4X^3$, combien de fois X^2 ? » Réponse, $4X$. On calcule alors $4X \cdot B = 4X^3 + 4X^2 + 4X$, et l'on soustrait ce résultat au polynôme A. On présente ces calculs de la manière suivante :

$$\begin{array}{cccc|l} 4X^3 & -2X^2 & & +1 & X^2+X+1 \\ 4X^3 & +4X^2 & +4X & & 4X \\ \hline & -6X^2 & -4X & +1 & \end{array}$$

On recommence avec « en $-6X^2$, combien de fois X^2 ? », réponse -6 :

$$\begin{array}{cccc|l} 4X^3 & -2X^2 & & +1 & X^2+X+1 \\ 4X^3 & +4X^2 & +4X & & 4X-6 \\ \hline & -6X^2 & -4X & +1 & \\ & -6X^2 & -6X & -6 & \\ \hline & & 2X & +7 & \end{array}$$

C'est terminé : lorsque l'on obtient à gauche un polynôme de degré inférieur à celui de B, c'est le reste, ici $R = 2X + 7$. Le quotient est $Q = 4X - 6$. On peut vérifier directement que $A = BQ + R$.

voir les exercices 364, 375

Exemple 11.4 – Dans ce qui suit, les divisions euclidiennes vont être d'une grande utilité, mais pour l'instant vous vous demandez peut-être quel intérêt on pourrait bien avoir à diviser des polynômes. Voici alors une petite astuce de calcul qui les fait intervenir. Soit

$$j = \frac{-1 + i\sqrt{3}}{2}.$$

C'est une solution de $X^2 + X + 1 = 0$, c'est-à-dire que $j^2 + j + 1 = 0$ (proposition 2.16). Combien vaut $4j^3 - 2j^2 + 1$? Si l'on commence par développer $(\frac{-1+i\sqrt{3}}{2})^3$ de manière naïve, on va perdre pas mal de temps. Alors que nous venons de démontrer que

$$4X^3 - 2X^2 + 1 = (4X - 6)(X^2 + X + 1) + 2X + 7,$$

ce qui donne en évaluant en $X = j$ la réponse $4j^3 - 2j^2 + 1 = 2j + 7 = 6 + i\sqrt{3}$.

Notez bien que la division euclidienne ne fait intervenir que des nombres entiers (et aucune $\sqrt{3}$), et qu'elle s'effectue très vite, avec l'habitude. Vous aviez peut-être réussi à calculer $4j^3 - 2j^2 + 1$ rapidement en écrivant $j^2 = -1 - j$, donc $j^3 = -j - j^2 = 1$. Bravo, mais la méthode de la division euclidienne ne fait rien d'autre que d'organiser ces calculs. Avec des nombres encore plus compliqués que ce j, il devient très difficile de trouver des astuces au coup d'œil.

Définition 11.5 – Soit $P \in \mathbb{K}[X]$ un polynôme, et soit $r \in \mathbb{K}$. On dit que r est une *racine* de P lorsque $P(r) = 0$. (Parfois on dit que r est une solution de P, et parfois on dit (assez curieusement, d'ailleurs) que r est un zéro de P.)

Proposition 11.6 – *Le nombre $r \in \mathbb{K}$ est une racine de P si et seulement si le polynôme $X - r$ divise P dans $\mathbb{K}[X]$.*

Démonstration. On écrit la division euclidienne de P par $X - r$:

$$P = (X - r)Q + R.$$

Ici le degré de R doit être < 1, donc R est de degré 0 (on dit que c'est une « constante »). En faisant $X = r$, ceci devient $P(r) = R$, donc finalement

$$P = (X - r)Q + P(r).$$

Il est alors clair que $P(r) = 0 \Leftrightarrow (X - r) \mid P$. □

La démonstration indique clairement que pour trouver explicitement le polynôme $\frac{P}{X-r}$, le plus simple est d'effectuer une division euclidienne.

Exemple 11.7 – Soit $P = 5X^2 - 15X + 10$. Ce polynôme a deux racines, à savoir 1 et 2. Il doit donc être divisible par $X - 1$ en particulier, et en faisant la division euclidienne on obtient

$$5X^2 - 15X + 10 = (X - 1)(5X - 10) = 5(X - 1)(X - 2),$$

ce qui confirme que P est également divisible par $X - 2$.

D'une manière générale, si P a une racine r_1 on peut écrire $P = (X - r_1)Q_1$, et si Q_1 a une racine r_2 on peut écrire $Q_1 = (X - r_2)Q_2$ donc $P = (X - r_1)(X - r_2)Q_2$; si Q_2 a une racine r_3 on aboutit à $P = (X - r_1)(X - r_2)(X - r_3)Q_3 \ldots$ Peut-on continuer comme ça indéfiniment ? En d'autres termes, est-ce que chaque polynôme de $\mathbb{K}[X]$ va toujours posséder au moins une racine dans $\mathbb{K}$?

La réponse est non, tout d'abord parce que les polynômes constants, de la forme $P(X) = c$, avec $c \in \mathbb{K}$, n'ont aucune racine si $c \neq 0$. Si maintenant $\deg P = n \geq 1$, on observe que les degrés successifs de $Q_1, Q_2, \ldots$, ne font que diminuer, donc si l'on peut trouver n racines successivement comme ci-dessus le polynôme Q_n sera de degré 0 donc constant et non-nul. On ne peut alors pas continuer avec Q_n.

Au passage nous avons presque démontré le résultat suivant :

Proposition 11.8 – *Un polynôme de degré n ne possède pas plus de n racines distinctes.*

Démonstration. Supposons en effet que l'on ait $n + 1$ racines distinctes, disons $r_1, r_2, \ldots, r_{n+1}$. On commence par écrire $P = (X - r_1)Q_1$ comme ci-dessus. Ensuite, puisque $P(r_2) = 0$, on écrit

$$P(r_2) = (r_2 - r_1)Q_1(r_2) = 0,$$

et comme $r_1 \neq r_2$ par hypothèse, on doit bien avoir $Q_1(r_2) = 0$. On peut donc factoriser et obtenir $Q_1 = (X - r_2)Q_2$. On recommence avec Q_2, et ainsi de suite on aboutit à

$$P = (X - r_1)(X - r_2) \cdots (X - r_{n+1})Q_{n+1}.$$

Cette dernière égalité est absurde puisque le membre de droite a un degré $\geq n + 1$. □

Mais il n'y a pas que les polynômes constants qui n'ont pas de racines. L'exemple le plus fameux est $P = X^2 + 1 \in \mathbb{R}[X]$ et qui ne possède pas de racine dans $\mathbb{R}$, puisque le carré d'un réel est toujours positif et ne saurait valoir -1. Dans le même ordre d'idées, le polynôme $Q = X^2 - 2 \in \mathbb{Q}[X]$ n'a pas de racine dans $\mathbb{Q}$ d'après la proposition 2.1.

Dans un cas comme dans l'autre, on peut considérer ces polynômes comme des éléments de $\mathbb{C}[X]$, et ils ont bien sûr des racines dans $\mathbb{C}$. Rappelez-vous qu'en parlant de $\mathbb{C}$ nous avions prédit que nous gagnerions bien plus que des racines carrées supplémentaires en travaillant avec les complexes. Le théorème suivant affirme en effet que toute équation polynomiale $P(z) = 0$ a une solution dans $\mathbb{C}$!

Théorème 11.9 (Théorème fondamental de l'algèbre) – *Tout polynôme de degré ≥ 1 dans $\mathbb{C}[X]$ possède une racine dans $\mathbb{C}$.*

On dit que « $\mathbb{C}$ est algébriquement clos ». Voir le théorème B.16, en appendice, pour une démonstration.

Corollaire 11.10 – *Tout polynôme de $\mathbb{C}[X]$ de degré n peut s'écrire de manière unique*

$$P = \lambda(X - r_1)(X - r_2) \cdots (X - r_n).$$

Démonstration. L'existence d'une telle écriture est claire à ce stade, mais comme nous énonçons l'unicité aussi on va prudemment faire une récurrence. Supposons le résultat vrai pour les polynômes de degré $n - 1$ (pour $n = 0$ c'est évident). Soit P de degré n.

On peut trouver une racine r_1 pour P d'après le théorème, donc $P = (X - r_1)Q$. Par récurrence on sait que $Q = \lambda(X - r_2)(X - r_3) \cdots (X - r_n)$, d'où l'écriture annoncée pour P.

Vérifions qu'elle est unique. Si

$$P = \mu(X - y_1)(X - y_2) \cdots (X - y_m),$$

on sait déjà que $m = n = \deg P$ et que $\mu = \lambda =$ le coefficient de X^n dans P. Comme $P(r_1) = 0$, on peut écrire

$$\mu(r_1 - y_1)(r_1 - y_2) \cdots (r_1 - y_n) = 0,$$

donc $r_1 - y_i = 0$ pour un certain indice i ; quitte à renuméroter, on peut supposer que $i = 1$, donc $y_1 = r_1$.

Le quotient dans la division de P par $(X - r_1)$, qui est uniquement déterminé, peut donc être calculé de deux façons différentes, ce qui donne l'égalité

$$Q = \lambda(X - y_2)(X - y_3) \cdots (X - y_n).$$

Par récurrence, on sait que cette écriture est unique, c'est-à-dire que (quitte à renuméroter) on a $x_i = y_i$ pour tous les indices i. □

La situation pour les polynômes de $\mathbb{R}[X]$ est à peine plus compliquée. Faisons une remarque simple :

Lemme 11.11 – *Soit $P \in \mathbb{R}[X]$, et soit $r \in \mathbb{C}$ une racine de P. Alors le nombre conjugué $\bar{r}$ est également racine de P.*

Démonstration. Si $P(r) = 0$, on a aussi $\overline{P(r)} = \bar{0} = 0$. Mais comme

$$P(r) = a_0 + a_1 r + \cdots + a_n r^n$$

avec $\overline{a_i} = a_i$ (puisque $a_i \in \mathbb{R}$), on constate que

$$\overline{P(r)} = a_0 + a_1 \bar{r} + \cdots + a_n \bar{r}^n = P(\bar{r}) = 0. \quad \square$$

Proposition 11.12 – *Soit P un polynôme de $\mathbb{R}[X]$. On peut écrire de manière unique*

$$P = \lambda(X - x_1)(X - x_2) \cdots (X - x_\ell)Q_1 Q_2 \cdots Q_j$$

avec $x_k \in \mathbb{R}$ et Q_k un polynôme de degré 2 dans $\mathbb{R}[X]$ sans racine réelle.

Démonstration. D'après le corollaire, on peut écrire

$$P = \lambda(X - r_1)(X - r_2) \cdots (X - r_n),$$

avec $r_k \in \mathbb{C}$. Ceux de ces nombres qui sont en fait dans $\mathbb{R}$ vont être notés $x_1, x_2, \ldots x_\ell$. Les autres racines r_k qui ne sont pas réelles sont regroupées en paires : en effet d'après le lemme, si r_k est une racine de P, alors $\overline{r_k}$ aussi (et $\overline{r_k} \neq r_k$ dans ce cas). Le facteur

$$(X - r_k)(X - \overline{r_k}) = X^2 - 2\mathfrak{Re}(r_k)X + |r_k|^2$$

est un polynôme de $\mathbb{R}[X]$ de degré 2, sans racine réelle. La proposition s'en déduit. voir les exercices 401, 412, 423

L'unicité est laissée en exercice. □

Que peut-on dire de l'ensemble des diviseurs d'un polynôme P donné ? Tout d'abord, notons que si $P = QR$, on peut écrire pour tout scalaire $\lambda \neq 0$ que $P = (\lambda Q)(\frac{1}{\lambda} R)$; en d'autres termes, si Q divise P, alors λQ divise aussi P, et en particulier P a un nombre infini de diviseurs.

Pour éviter cette complication inutile, nous dirons qu'un polynôme $a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$ est *unitaire* si $a_n = 1$. Une meilleure question est donc : que peut-on dire de l'ensemble des diviseurs unitaires d'un polynôme donné ? Notons que, à regarder les définitions, il n'est pas *a priori* évident de savoir s'il en existe un nombre fini ou non.

Grâce au corollaire 11.10 cependant, on va pouvoir étudier facilement l'ensemble des diviseurs d'un polynôme complexe. Introduisons la notation suivante : pour un nombre complexe z et un polynôme $P \in \mathbb{C}[X]$, le nombre $m_z(P)$ est le plus grand entier tel que $(X - z)^{m_z(P)}$ divise P. On va l'appeler la *multiplicité* de z comme racine de P. On remarque que $m_z(P) > 0$ si et seulement si z est racine de P, ce qui n'arrive que pour un nombre fini de nombres z. Le corollaire 11.10 peut s'écrire

$$P = \lambda \prod_{z \in \mathbb{C}} (X - z)^{m_z(P)},$$

sachant que ce produit ne comporte qu'un nombre fini de termes (les autres sont égaux à 1). Le lemme suivant est alors évident.

LEMME 11.13 – *Les multiplicités ont les propriétés suivantes.*

1. $m_z(P_1 P_2) = m_z(P_1) + m_z(P_2)$.
2. Q *divise* P *si et seulement si* $m_z(Q) \leq m_z(P)$ *pour tout nombre complexe* z.
3. P *ne possède qu'un nombre fini de diviseurs unitaires.*

Démonstration. La démonstration (très simple) des points (1) et (2) est laissée en exercice. Pour le (3) notons qu'un polynôme unitaire Q s'écrit

$$Q = \prod_{z \in \mathbb{C}} (X - z)^{m_z(Q)},$$

et lorsque Q divise P le nombre entier $m_z(Q)$ doit vérifier $0 \leq m_z(Q) \leq m_z(P)$ d'après le point (2). Il n'y a qu'un nombre fini de nombres complexes z pour lesquels $m_z(P) \neq 0$, donc au total il n'y a qu'un nombre fini de choix pour Q. □

On peut alors poser la définition suivante.

DÉFINITION 11.14 – Soit P et Q deux polynômes de $\mathbb{C}[X]$. Le polynôme pgcd(P, Q) est par définition

$$\mathrm{pgcd}(P, Q) = \prod_{z \in \mathbb{C}} (X - z)^{\min(m_z(P), m_z(Q))}.$$

LEMME 11.15 – *Le polynôme* pgcd(P, Q) *divise* P *et divise* Q. *De plus si* D *est un polynôme qui divise* P *et* Q, *alors* D *divise également* pgcd(P, Q). *En particulier* pgcd(P, Q) *est l'unique diviseur unitaire de* P *et de* Q *de degré maximal.*

On comprend donc pourquoi pgcd(P, Q) est appelé le plus grand diviseur commun de P et Q.

Démonstration. D'après le (2) du lemme précédent, il est clair que si $D \mid P$ et $D \mid Q$, alors $D \mid \mathrm{pgcd}(P, Q)$, et donc que $\deg(D) \leq \deg(\mathrm{pgcd}(P, Q))$.

Vérifions l'unicité. Soit D un diviseur unitaire de P et de Q dont le degré est maximal. On a $D \mid \mathrm{pgcd}(P, Q)$ donc $\deg(D) \leq \deg(\mathrm{pgcd}(P, Q))$ d'où par maximalité $\deg(D) = \mathrm{pgcd}(P, Q)$. Par suite $D = \mathrm{pgcd}(P, Q)$ puisqu'ils sont tous les deux unitaires. □

Cette approche des pgcds a pas mal de défauts. Pour commencer, il est difficile de calculer pgcd(P, Q) par la définition ci-dessus : il faut d'abord factoriser entièrement P et Q ! Ensuite, si $\mathbb{K}$ n'est pas $\mathbb{C}$, on ne sait rien dire. Vous arriverez certainement à traiter le cas $\mathbb{K} = \mathbb{R}$ à l'aide de la proposition 11.12 (en lieu et place du corollaire 11.10), mais pour $\mathbb{K} = \mathbb{Q}$ on est dans une impasse.

Dans la suite du chapitre on va indiquer une toute autre méthode, plus générale et entraînant des calculs assez faciles. Par contre les définitions sont moins directes.

– Fin de la première lecture –

DÉFINITION 11.16 – Soient A et B deux polynômes (ou deux nombres entiers). L'ensemble des diviseurs communs à A et à B est noté $\mathrm{div}(A, B)$.

Notez que $\mathrm{div}(A, 0)$ est l'ensemble des diviseurs de A (tout polynôme P divise le polynôme nul, puisque $0 = 0 \times P$).

LEMME 11.17 – *Soient A et $B \neq 0$ des polynômes (ou des nombres entiers). Écrivons la division euclidienne $A = BQ + R$. Alors*

$$\mathrm{div}(A, B) = \mathrm{div}(B, R).$$

Démonstration. Si D divise A et B, alors il divise $R = A - BQ$ (en effet si $A = A'D$ et $B = B'D$ alors $R = (A' - B'Q)D$). Réciproquement si D divise R et B, alors il divise A, par le même raisonnement. Donc les diviseurs à considérer pour la paire (A, B) sont les mêmes que pour la paire (B, R). □

Pourquoi est-ce utile ? Tout simplement parce qu'en passant à (B, R), les degrés (ou les nombres) sont plus petits. On peut ensuite recommencer avec (B, R), et recommencer encore, et on va finir par obtenir une paire de la forme $(P, 0)$: en effet tant que le deuxième terme n'est pas nul, on fait une nouvelle division euclidienne, et on obtient un nouveau terme strictement plus petit. En fait on a :

PROPOSITION 11.18 – *Soient A et B des polynômes.*

1. *Il existe un unique polynôme unitaire P tel que*

$$\mathrm{div}(A, B) = \mathrm{div}(P, 0).$$

On le note $\mathrm{pgcd}(A, B)$.

2. *Si D divise A et B, alors D divise également leur pgcd.*
3. *Le polynôme* $\mathrm{pgcd}(A, B)$ *est également caractérisé comme l'unique diviseur unitaire commun à A et à B dont le degré est maximal.*
4. *Si on effectue une division euclidienne $A = BQ + R$, alors*

$$\mathrm{pgcd}(A, B) = \mathrm{pgcd}(B, R).$$

Cette définition du pgcd est donc cohérente avec la définition 11.14 lorsque $\mathbb{K} = \mathbb{C}$ (à cause du point (3)).

Démonstration. (1) Nous venons d'expliquer comment, en appliquant le lemme précédent suffisamment souvent, on trouve un polynôme P tel que $\mathrm{div}(A, B) = \mathrm{div}(P, 0)$; on peut supposer P unitaire. Montrons l'unicité. Si $\mathrm{div}(P, 0) = \mathrm{div}(P', 0)$, alors comme $P \mid P$ on a aussi $P \mid P'$; et réciproquement comme $P' \mid P'$ on a $P' \mid P$. Finalement P et P' se divisent l'un l'autre, et sont unitaires, donc $P = P'$. Le polynôme P est bien unique, et on peut le noter $\mathrm{pgcd}(A, B)$.

(2) L'assertion sur D n'est qu'une traduction de l'égalité entre $\mathrm{div}(A, B)$ et $\mathrm{div}(P, 0)$.

(3) Soit $D \in \mathrm{div}(A, B)$ de degré maximal. On a $D \mid \mathrm{pgcd}(A, B)$, donc $\deg(D) \leq \deg(\mathrm{pgcd}(A, B))$ d'où par maximalité $\deg(D) = \mathrm{pgcd}(A, B)$. Par suite, on a bien $D = \mathrm{pgcd}(A, B)$ puisqu'ils sont tous les deux unitaires.

(4) D'après le lemme précédent, on a $\mathrm{div}(A, B) = \mathrm{div}(B, R)$, donc c'est évident. □

Avant de donner des exemples, remarquons que la situation avec les nombres entiers est exactement similaire. En fait on a :

PROPOSITION 11.19 – *Soient a et b des nombres entiers.*

1. *Il existe un unique entier positif p tel que*

$$\mathrm{div}(a, b) = \mathrm{div}(p, 0).$$

On le note $\mathrm{pgcd}(a, b)$.

2. *Si d divise a et b, alors d divise également leur pgcd.*
3. *Le nombre $\mathrm{pgcd}(a, b)$ est également caractérisé comme le plus grand diviseur commun à a et à b (!).*
4. *Si on effectue une division euclidienne $a = bq + r$, alors*

$$\mathrm{pgcd}(a, b) = \mathrm{pgcd}(b, r).$$

La démonstration est la même. Voyons des exemples.

EXEMPLE 11.20 – Commençons par des nombres entiers, disons par exemple $a = 77$ et $b = 91$. On écrit

$$91 = 77 \times 1 + 14,$$

donc $\mathrm{pgcd}(91, 77) = \mathrm{pgcd}(77, 14)$. Puis

$$77 = 14 \times 5 + 7,$$

donc $\mathrm{pgcd}(77, 14) = \mathrm{pgcd}(14, 7)$. Finalement

$$14 = 7 \times 2 + 0,$$

donc $\mathrm{pgcd}(14, 7) = \mathrm{pgcd}(7, 0) = 7$. Au total $\mathrm{pgcd}(77, 91) = 7$.

EXEMPLE 11.21 – Prenons $A = X^3 + 7X^2 + 2X + 14$ et $B = X^4 + 4X^2 + 4$, dans $\mathbb{R}[X]$. On calcule

$$B = (X - 7) \cdot A + (51X^2 + 102),$$

donc $\mathrm{pgcd}(B, A) = \mathrm{pgcd}(A, R)$ avec $R = 51X^2 + 102$ (on passe les détails du calcul de la division euclidienne). Puis on effectue

$$A = \left(\frac{1}{51}X + \frac{7}{51}\right) \cdot R + 0,$$

donc $\mathrm{pgcd}(A, R) = \mathrm{pgcd}(R, 0)$. Attention, comme le pgcd est un polynôme unitaire par définition, ici la réponse n'est pas $R = 51(X^2 + 2)$ mais $X^2 + 2$: on divise simplement par le coefficient du terme en X^2. Finalement $\mathrm{pgcd}(A, B) = X^2 + 2$.

voir les exercices 379, 380

À propos, la méthode de calcul du pgcd que nous venons d'appliquer, à base de divisions euclidiennes successives, s'appelle *l'algorithme d'Euclide*.

C'est le suivant :

Théorème 11.22 – *Soient a et b deux nombres entiers. Alors il existe deux nombres u et v tels que*

$$au + bv = \operatorname{pgcd}(a,b).$$

Soient A *et* B *deux polynômes de* $\mathbb{K}[X]$*. Alors il existe deux polynômes* $U, V \in \mathbb{K}[X]$ *tels que*

$$AU + BV = \operatorname{pgcd}(A,B).$$

Démonstration. Dans l'algorithme d'Euclide, on passe d'une paire à la suivante en ajoutant des multiples de a et des multiples de b. □

Exemple 11.23 – Revenons à l'exemple 11.20, avec $a = 77$ et $b = 91$; on a vu que $d = \operatorname{pgcd}(a,b) = 7$.

Reprenons les divisions euclidiennes que nous avons faites, en exprimant systématiquement les restes en fonction de a et b. Nous sommes partis de

$$91 = 77 \times 1 + 14 \quad \text{donc} \quad 14 = b - a.$$

Puis nous avons effectué

$$77 = 14 \times 5 + 7 \quad \text{donc} \quad 7 = a - 5(b-a) = 6a - 5b.$$

Enfin la dernière division nous a montré que le pgcd était bien $d = 7$. On a donc $d = 6a - 5b$, ce qui est bien la formule annoncée avec $u = 6$ et $v = -5$.

Exemple 11.24 – Cette fois, reprenons l'exemple 11.21. La première division était

$$B = (X-7)\cdot A + 51D,$$

avec $D = \operatorname{pgcd}(A,B) = X^2 + 2$. On a donc bien

$$D = -\frac{1}{51}(X-7)A + \frac{1}{51}B,$$

voir les exercices 387, 370

qui est la forme annoncée avec $U = -\frac{1}{51}(X-7)$ et $V = \frac{1}{51}$.

Un nombre ou un polynôme va être appelé premier lorsqu'il n'a aucun diviseur à part ceux qui sont évidents. Précisons :

Définition 11.25 – Soit $p \in \mathbb{Z}$ un nombre $\neq \pm 1$. On dit que p est *premier* lorsque la seule façon d'obtenir une factorisation $p = ab$ (avec $a, b \in \mathbb{Z}$) est de prendre $a = \pm 1$ ou $b = \pm 1$.

Soit $P \in \mathbb{K}[X]$ un polynôme de degré ≥ 1. On dit que P est *premier*, ou plus souvent *irréductible*, lorsque la seule façon d'obtenir une factorisation $P = AB$ (avec $A, B \in \mathbb{K}[X]$) est de prendre A constant ou B constant.

Exemple 11.26 – Un nombre p est donc premier si et seulement si la liste complète de ses diviseurs est $\{1, -1, p, -p\}$. Les nombres 17, 71, −277, −733 et 953 sont ainsi premiers. On adopte souvent la convention de ne parler que des nombres premiers positifs, de sorte que leur liste commence par 2, 3, 5, 7, 11, 13 ...

Exemple 11.27 – Un polynôme de degré 1 est toujours irréductible (premier). Pour un polynôme $P \in \mathbb{K}[X]$ de degré 2, la situation est encore assez simple. Si $P = AB$ et si ni A, ni B n'est constant, alors $\deg A = \deg B = 1$. Comme un polynôme de degré 1 possède toujours une racine dans $\mathbb{K}$, le polynôme P en possède aussi une. Mais la réciproque est vraie : si $P(r) = 0$, alors en posant $A = X - r$ on peut écrire $P = AB$ avec B de degré 1 (proposition 11.6).

On retiendra qu'un polynôme de $\mathbb{K}[X]$ de degré 2 est irréductible si et seulement si il ne possède pas de racine dans $\mathbb{K}$. Ça reste vrai pour un polynôme de degré 3 (vérifiez-le).

Par exemple $P = X^2 + 1$ est irréductible dans $\mathbb{R}[X]$. Par contre dans $\mathbb{C}[X]$ on a $P = (X - i)(X + i)$.

Proposition 11.28 – *Un polynôme de $\mathbb{C}[X]$ est irréductible si et seulement s'il est de degré 1. Un polynôme de $\mathbb{R}[X]$ est irréductible si et seulement s'il est ou bien de degré 1, ou bien de degré 2 sans racine dans $\mathbb{R}$.*

Démonstration. Commençons par $\mathbb{C}$. Les choses sont simples : nous venons de remarquer qu'un polynôme de degré 1 est toujours irréductible, et réciproquement le corollaire 11.10 montre qu'un polynôme de degre > 1 peut se factoriser et n'est donc pas premier.

C'est presque la même chose sur $\mathbb{R}$. Dans l'exemple précédent nous avons montré que les polynômes en question sont effectivement irréductibles, et que les polynômes de degré 2 irréductibles n'ont pas de racine dans $\mathbb{R}$. Ensuite, c'est la proposition 11.12 qui montre que tout polynôme de degré 3 ou plus se factorise comme produits de tels polynômes, et donc n'est pas irréductible. □

Le théorème de Bézout va permettre de démontrer des choses sur les nombres premiers et les polynômes irréductibles qui paraissent intuitives, mais qu'on ne saurait pas prouver autrement. Voici le meilleur exemple.

Lemme 11.29 (Lemme de Gauss) – *Soit* P *un premier. Si* $P \mid AB$, *alors* $P \mid A$ *ou* $P \mid B$.

Ce résultat est valable pour les entiers comme pour les polynômes.

Démonstration. Supposons que P ne divise pas A, et montrons qu'il divise alors B. Notons que l'on a $\operatorname{pgcd}(P, A) = 1$, et donc $UA + PV = 1$ par Bézout. Par hypothèse on a $AB = PR$, donc $UAB = UPR$, et comme $UA = 1 - PV$, on en tire

$$(UA)B = B - PVB = UPR .$$

Ceci montre que $B = P(VB + UR)$ est bien divisible par P. □

Nous pouvons finalement apporter une réponse à la question soulevée à la fin du chapitre 2 (cf corollaire 2.25) : quand est-ce que $\mathbb{Z}/N\mathbb{Z}$ est un corps ?

Proposition 11.30 – *Soit p un entier ≥ 2. Alors $\mathbb{Z}/p\mathbb{Z}$ est un corps si et seulement si p est un nombre premier.*

Démonstration. Si p n'est pas premier, on a $p = ab$ avec a et b des nombres qui ne sont pas divisibles par p. En réduisant modulo p, on obtient

$$\overline{a}\,\overline{b} = \overline{p} = \overline{0} ,$$

avec $\overline{a} \neq \overline{0}$ et $\overline{b} \neq \overline{0}$. Il est donc impossible que $\overline{a}$ ait un inverse (argumenter comme dans l'exemple 2.27).

Réciproquement, supposons que p soit premier, et soit $\overline{a} \neq \overline{0}$ un élément de $\mathbb{Z}/p\mathbb{Z}$. L'entier a est alors premier avec p, c'est-à-dire que $\operatorname{pgcd}(p, a) = 1$ puisque l'on suppose que p ne divise pas a. Par Bézout, on a $au + pv = 1$, et en réduisant modulo p on a

$$\overline{a}\,\overline{u} + \overline{p}\,\overline{v} = \overline{a}\,\overline{u} = \overline{1}$$

(étant donné que $\overline{p} = \overline{0}$). C'est donc bien que $(\overline{a})^{-1} = \overline{u}$. Tout nombre non-nul de $\mathbb{Z}/p\mathbb{Z}$ possède un inverse, et on a bien affaire à un corps. □

Le théorème suivant généralise le corollaire 11.10 et la proposition 11.12.

Théorème 11.31 – *Soit* $P \in \mathbb{K}[X]$. *On peut écrire de manière unique*

$$P = \lambda P_1 P_2 \cdots P_k \,,$$

où chaque P_i *est irréductible et unitaire.*

(Lisez la démonstration pour plus de précisions sur ce que « de manière unique » signifie.)

Démonstration. Montrons l'existence de cette écriture, par récurrence sur le degré de P (c'est évident si $\deg(P) = 0$). On peut supposer que P est unitaire. Si P est lui-même irréductible, alors il n'y a rien à dire. Dans le cas contraire, on écrit $P = QR$ avec $\deg(Q) < \deg(P)$ et également $\deg(R) < \deg(P)$. Par récurrence on peut factoriser Q et R en produit d'irréductibles, donc P aussi.

Montrons l'unicité (c'est plus fin). On doit donc montrer que si on a deux écritures

$$\lambda P_1 P_2 \cdots P_k = \mu Q_1 Q_2 \cdots Q_\ell \,, \qquad (*)$$

avec chaque P_i et chaque Q_i unitaire et irréductible, alors $\lambda = \mu$, $k = \ell$, et les polynômes P_i sont les mêmes que les Q_i. On procède par récurrence sur le degré des deux membres de (*) (les choses sont évidentes si le degré est 0).

Tout d'abord en regardant le coefficient de plus haut degré, on voit de suite que $\lambda = \mu$. Ensuite, on constate que P_1 divise le produit $Q_1 Q_2 \cdots Q_\ell$. En conséquence du lemme de Gauss, P_1 doit diviser l'un des Q_i, disons Q_1 pour simplifier les notations. Par suite on a $P_1 = Q_1$, puisqu'il n'y a qu'un seul diviseur unitaire $\neq 1$ du polynôme irréductible Q_1, à savoir Q_1 lui-même. On simplifie maintenant (*) par $\lambda P_1 = \lambda Q_1$ pour obtenir

$$P_2 P_3 \cdots P_k = Q_2 Q_3 \cdots Q_\ell \,. \qquad (**)$$

L'égalité (**) est de degré plus petit que (*), donc par récurrence on sait que $k = \ell$ et que les polynômes $P_2, \ldots, P_k$ sont les mêmes que les polynômes $Q_2, \ldots, Q_\ell$. Ceci termine la démonstration. □

Ce résultat est extrêmement fort. C'est surtout en deuxième année que vous en explorerez les conséquences plus sérieusement. En attendant, prenez le temps de vérifier la chose suivante : non seulement le théorème généralise le corollaire 11.10 et la proposition 11.12 comme nous l'annoncions, mais il démontre aussi pour $\mathbb{K}$ quelconque qu'un polynôme ne possède qu'un nombre fini de diviseurs unitaires ; plus généralement vous pouvez maintenant, pour $\mathbb{K}$ quelconque, montrer une version du lemme 11.13 (il faut commencer par définir les « multiplicités » $m_Q(P)$ pour Q irréductible).

Par ailleurs, ce théorème existe pour les nombres entiers, essentiellement avec la même démonstration, et vous le connaissez probablement déjà :

Théorème 11.32 – *Soit* $n \in \mathbb{Z}$. *On suppose que* $n \notin \{-1, 0, 1\}$. *Alors on peut écrire de manière unique*

$$n = \pm p_1 p_2 \cdots p_k \,,$$

où chaque p_i *est un nombre premier positif.*

Chapitre 12

Matrices

On rappelle que la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$.

Une *matrice* n'est rien d'autre qu'un tableau de nombres. Si l'on s'intéresse mathématiquement aux tableaux, c'est toujours de près ou de loin parce qu'ils interviennent dans les *systèmes linéaires*, c'est-à-dire les équations du genre

$$\left\{\begin{array}{ccccccc} 3x & - & 7y & + & 9z & = & 1 \\ 2x & + & y & & & = & 3 \end{array}\right. .$$

Ici on pourra associer à ce système la matrice

$$\begin{pmatrix} 3 & -7 & 9 & 1 \\ 2 & 1 & 0 & 3 \end{pmatrix}.$$

Il est clair que toutes les informations concernant le système sont contenues dans cette matrice, mais de plus on va voir (et vous en avez sans doute eu un aperçu au lycée) que la résolution du système gagne même en clarté lorsqu'elle est faite à partir de manipulations sur la matrice. L'observation-clef que nous allons voir tout de suite est que les matrices peuvent être *multipliées* entre elles. Il en résulte des notations simples et puissantes.

Dans les chapitres suivants, nous montrerons que les systèmes linéaires, sous des formes plus sophistiquées, sont omniprésents en mathématiques, à un point qui devrait vous surprendre. Dans ce chapitre le but est simplement de se familiariser avec les matrices, et d'apprendre à résoudre les systèmes.

Commençons par les définitions.

Définition 12.1 – Une *matrice* de type $n \times m$ à coefficients dans $\mathbb{K}$ est un tableau d'éléments de $\mathbb{K}$ comprenant n lignes et m colonnes. L'ensemble des matrices $n \times m$ est noté $M_{n,m}(\mathbb{K})$. On utilise parfois la notation $M_n(\mathbb{K})$ au lieu de $M_{n,n}(\mathbb{K})$, dans le cas des matrices « carrées ».

Par exemple,

$$\begin{pmatrix} 3 & -7 & 9 & 1 \\ 2 & 1 & 0 & 3 \end{pmatrix} \in M_{2,4}(\mathbb{K}) \quad \text{et} \quad \begin{pmatrix} 2-3i & -4 \\ \sqrt{17} & \frac{2}{3} \end{pmatrix} \in M_2(\mathbb{C}).$$

Les éléments de $M_{1,n}(\mathbb{K})$ sont appelés *matrices-lignes* et ceux de $M_{n,1}(\mathbb{K})$ sont les *matrices-colonnes*.

Notez bien la convention suivante. À partir de maintenant, nous allons identifier les matrices-*colonnes* avec les vecteurs de $\mathbb{K}^n$; c'est-à-dire que l'on identifie $\mathbb{K}^n$ et $M_{n,1}(\mathbb{K})$. On se permet donc d'écrire

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{K}^n .$$

On fait ce choix particulier dans le but de simplifier (énormément) certaines formules qui vont apparaître dans la suite. Dans d'autres chapitres de ce livre, vous l'avez sans doute déjà remarqué, les éléments de $\mathbb{K}^n$ sont notés $(x_1, \ldots, x_n)$, donc en ligne : il faut voir ça comme une notation que l'on s'autorise dès qu'il n'y a pas d'ambiguïté, dans le but d'économiser la place. Mais dès lors qu'il y a des matrices en vue, et des opérations sur ces matrices, les vecteurs sont des *colonnes*.

On utilise parfois la notation

$$(a_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le m}}$$

ou plus simplement $(a_{ij})_{i,j}$ lorsque n et m sont entendus, pour désigner la matrice de $M_{n,m}(\mathbb{K})$ donc le coefficient sur la ligne i, dans la colonne j, est le nombre a_{ij}. Par exemple dans le cas de matrices 2×3, la matrice $(a_{ij})_{i,j}$ est

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}.$$

Autre exemple, avec des matrices 3×3, la « matrice donnée par $(\cos(i)\sin(j))_{i,j}$ » est

$$\begin{pmatrix} \cos(1)\sin(1) & \cos(1)\sin(2) & \cos(1)\sin(3) \\ \cos(2)\sin(1) & \cos(2)\sin(2) & \cos(2)\sin(3) \\ \cos(3)\sin(1) & \cos(3)\sin(2) & \cos(3)\sin(3) \end{pmatrix}.$$

Définition 12.2 – Si $A = (a_{ij})_{i,j}$ est une matrice $n \times m$, alors sa *transposée* est la matrice ${}^tA$, de type $m \times n$, donnée par

$${}^tA = (a_{ji})_{\substack{1 \le i \le m \\ 1 \le j \le n}} .$$

(Noter l'inversion de i et de j : sur la ligne i dans la colonne j, on trouve cette fois a_{ji}.)

Par exemple si A est la matrice 2×3 ci-dessus, alors

$${}^tA = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}.$$

Si B est la matrice 3×3 ci-dessus, alors

$${}^tB = \begin{pmatrix} \cos(1)\sin(1) & \cos(2)\sin(1) & \cos(3)\sin(1) \\ \cos(1)\sin(2) & \cos(2)\sin(2) & \cos(3)\sin(2) \\ \cos(1)\sin(3) & \cos(2)\sin(3) & \cos(3)\sin(3) \end{pmatrix}.$$

La transposée d'une matrice-ligne est une matrice-colonne, et vice-versa.

Pour commencer, on peut additionner deux matrices de même type « coefficient par coefficient » : par exemple

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}.$$

(En d'autres termes $(a_{ij})_{i,j} + (b_{ij})_{i,j} = (a_{ij} + b_{ij})_{i,j}$.)

On peut aussi multiplier une matrice par un « scalaire », c'est-à-dire un élément de $\mathbb{K}$, en multipliant tous les coefficients par ce nombre : par exemple

$$2 \cdot \begin{pmatrix} 1 & 3 \\ 9 & 27 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 18 & 54 \end{pmatrix}.$$

(En d'autres termes $\lambda(a_{ij})_{i,j} = (\lambda a_{ij})_{i,j}$.)

Multiplier deux matrices entre elles est plus compliqué. Commençons par un cas simple.

Définition 12.3 – Donnons-nous une matrice-ligne de type $1 \times m$, disons

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_m \end{pmatrix};$$

prenons également une matrice-colonne de type $m \times 1$, disons

$$B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Alors le produit AB est par définition la matrice 1×1 dont l'unique coefficient est

$$a_1 b_1 + a_2 b_2 + \cdots + a_m b_m.$$

(Si le nombre de coefficients de A n'était pas égal au nombre de coefficients de B, le produit AB ne serait pas défini.)

Par exemple

$$\begin{pmatrix} 4 & -1 & 10 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \times (-1) + (-1) \times 0 + 10 \times 1 \end{pmatrix} = \begin{pmatrix} 6 \end{pmatrix};$$

ou encore

$$\begin{pmatrix} 2 & -3 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x - 3y \end{pmatrix}.$$

On va identifier, chaque fois que c'est commode, les matrices 1×1 et les éléments de $\mathbb{K}$. Ceci étant, dans le cas général, on définit la multiplication de la manière suivante.

Définition 12.4 – Soit A une matrice $n \times m$, et soit B une matrice $m \times p$. Alors le produit AB est la matrice $n \times p$ dont le coefficient sur la ligne i, dans la colonne j, est le produit de la ligne i de A par la colonne j de B.

Exemple 12.5 – Admettons que l'on souhaite multiplier

$$A = \begin{pmatrix} 1 & -4 \\ 1 & 2 \\ -1 & 1 \end{pmatrix}$$

par

$$B = \begin{pmatrix} -3 & -14 & -5 & 1 \\ 1 & 0 & -19 & -7 \end{pmatrix}.$$

Le produit est bien défini puisque le nombre de colonnes de A est égal au nombre de lignes de B. Le produit AB va avoir autant de lignes que A et autant de colonnes que B ; ce sera donc une matrice 3×4.

Pour faire le calcul, il est utile de présenter les choses comme ci-dessous.

$$\begin{array}{cc} & \begin{pmatrix} -3 & -14 & -5 & 1 \\ 1 & 0 & -19 & -7 \end{pmatrix} \\ \begin{pmatrix} 1 & -4 \\ 1 & 2 \\ -1 & 1 \end{pmatrix} & \begin{matrix} & & 71 & \\ & & & \\ & & & \end{matrix} \end{array}$$

Pour calculer, par exemple, le coefficient sur la première ligne, dans la colonne 3, du produit AB, on regarde la ligne et la colonne correspondantes. Ensuite on les multiplie par la méthode déjà donnée, donc ici

$$\begin{pmatrix} 1 & -4 \end{pmatrix} \cdot \begin{pmatrix} -5 \\ -19 \end{pmatrix} = \begin{pmatrix} 1 \times (-5) + (-4) \times (-19) \end{pmatrix} = \begin{pmatrix} 71 \end{pmatrix}.$$

On procède de même pour tous les autres coefficients. Après un certain temps, le calcul terminé ressemble à ceci :

$$\begin{array}{cc} & \begin{pmatrix} -3 & -14 & -5 & 1 \\ 1 & 0 & -19 & -7 \end{pmatrix} \\ \begin{pmatrix} 1 & -4 \\ 1 & 2 \\ -1 & 1 \end{pmatrix} & \begin{pmatrix} -7 & -14 & 71 & 29 \\ -1 & -14 & -43 & -13 \\ 4 & 14 & -14 & -8 \end{pmatrix} \end{array}$$

La matrice en bas à droite est AB.

voir les exercices 1040, 1041, 1042, 1043

Exemple 12.6 – Voyons comment la multiplication des matrices permet d'écrire les systèmes. Considérons par exemple

$$\begin{cases} 2x + 6y = -9 \\ -4x + 3y = 1 \end{cases}.$$

Posons alors

$$A = \begin{pmatrix} 2 & 6 \\ -4 & 3 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} -9 \\ 1 \end{pmatrix}.$$

On calcule facilement que

$$AX = \begin{pmatrix} 2x + 6y \\ -4x + 3y \end{pmatrix}.$$

Ainsi le système de départ est équivalent à une seule égalité de matrices-colonnes, à savoir :

$$AX = B.$$

Cette notation très compacte fait ressortir le plus important : on a bien envie de « diviser par A » des deux côtés, puisqu'on cherche X. Nous allons voir rapidement si oui ou non on peut donner un sens précis à cette intuition.

Il faut s'habituer à multiplier les matrices relativement vite et sans se tromper.

Il y a deux matrices qui jouent des rôles particuliers dans les opérations arithmétiques. Tout d'abord la matrice dont tous les coefficients sont nuls : on la note simplement 0, quelle que soit sa taille. Dans le cas des matrices carrées de taille n, on a également la matrice « identité » ci-dessous :

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & & \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

On la note Id_n ou Id ou encore I.

Proposition 12.7 – *Les règles de calcul suivantes sont valables dans* $\mathrm{M}_n(\mathbb{K})$ *:*

(a) $A + B = B + A$,
(b) $0 + A = A$,
(c) $(A + B) + C = A + (B + C)$,
(d) $\forall A\ \exists(-A)$ *tel que* $A + (-A) = 0$,
(e) $A(B + C) = AB + AC$ *et* $(A + B)C = AC + BC$,
(f) $\mathrm{Id}\,A = A\,\mathrm{Id} = A$,
(g) $(AB)C = A(BC)$.

(En d'autres termes, $\mathrm{M}_n(\mathbb{K})$ *est un anneau, les rôles de 0 et 1 étant joués par les matrices nulle et identité. Cet anneau n'est pas commutatif.)*

De plus l'opération de multiplication par un scalaire vérifie

$$\lambda \cdot A = (\lambda \cdot \mathrm{Id})A = A(\lambda \cdot \mathrm{Id}).$$

La démonstration est facile et vous est laissée à titre d'exercice. Par exemple examinez bien l'égalité $\mathrm{Id}\,A = A\,\mathrm{Id} = A$. C'est à cause de cette règle que l'on écrit parfois 1 pour la matrice identité.

Cet énoncé signifie que les règles de calcul habituelles s'appliquent aux matrices, *sauf* les deux suivantes. D'abord la commutativité : on n'a pas toujours $AB = BA$, par exemple pour

$$A = \begin{pmatrix} 0 & 3 \\ -1 & -1 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} -2 & -1 \\ 2 & -1 \end{pmatrix},$$

on a

$$AB = \begin{pmatrix} 6 & -3 \\ 0 & 2 \end{pmatrix} \quad \text{et} \quad BA = \begin{pmatrix} 1 & -5 \\ 1 & 7 \end{pmatrix}.$$

L'autre chose à laquelle il faut s'habituer, c'est qu'on ne peut pas toujours obtenir « l'inverse » d'une matrice. En d'autres termes, étant donnée A, il n'existe pas toujours de matrice A^{-1} telle que $AA^{-1} = \mathrm{Id}$, même si $A \neq 0$ (rappelons que le rôle de 1 est joué par Id !). En effet il suffit de prendre

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix};$$

il est clair qu'un produit AB donne toujours une matrice dont la deuxième ligne est nulle, donc on n'obtiendra jamais l'identité. Ce phénomène est très important pour la suite, donc soyons plus précis.

Définition 12.8 (et Proposition) – Soit $A \in \mathrm{M}_n(\mathbb{K})$. On dit que A est *inversible* lorsqu'il existe une matrice $B \in \mathrm{M}_n(\mathbb{K})$ telle que

$$AB = \mathrm{Id} \quad \text{et} \quad BA = \mathrm{Id}.$$

Une telle matrice B, lorsqu'elle existe, est unique. On la note A^{-1} et on l'appelle *l'inverse* de A.

Démonstration. Nous devons montrer l'unicité de B. Supposons donc que $AB = BA = \mathrm{Id}$ d'une part, et $AC = CA = \mathrm{Id}$ d'autre part. En multipliant $AB = \mathrm{Id}$ par C à gauche, on obtient

$$C(AB) = (CA)B = \mathrm{Id}\,B = B = C\,\mathrm{Id} = C.$$

On a donc bien $B = C$. □

On vient de voir un exemple de matrice non-inversible. Dans le cas des matrices 2×2, on peut décrire les matrices inversibles très facilement :

Proposition 12.9 – *Soit*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Alors A est inversible si et seulement si $ad - bc \neq 0$. *Dans ce cas l'inverse est donnée par*

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

La quantité $ad - bc$ s'appelle le déterminant de A. Dans le chapitre du même nom, nous verrons comment étendre ceci aux matrices $n \times n$.

Démonstration. Soit

$$\tilde{A} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Alors

$$A\tilde{A} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}.$$

Si $ad - bc = 0$, alors $A\tilde{A} = 0$. Dans ce cas la matrice A ne peut pas être inversible : si elle l'était, on aurait $A^{-1}A\tilde{A} = \tilde{A} = 0$, donc $a = b = c = d = 0$, donc $A = 0$, et donc $AA^{-1} = 0 = \mathrm{Id}$, contradiction !

Supposons maintenant que $ad - bc \neq 0$. On a

$$A\tilde{A} = (ad - bc)\,\mathrm{Id},$$

donc en multipliant par le scalaire $\frac{1}{ad-bc}$, on obtient

$$A\left(\frac{1}{ad - bc}\tilde{A}\right) = \mathrm{Id}.$$

De la même manière, vous vérifierez que dans l'autre sens on a aussi

$$\left(\frac{1}{ad - bc}\tilde{A}\right)A = \mathrm{Id},$$

voir l'exercice 1045

donc l'inverse de A existe et c'est bien $\frac{1}{ad-bc}\tilde{A}$, comme annoncé dans la proposition. □

Exemple 12.10 – Revenons à l'exemple 12.6. On y trouvait la matrice

$$A = \begin{pmatrix} 2 & 6 \\ -4 & 3 \end{pmatrix}.$$

Son déterminant est $2 \times 3 + 4 \times 6 = 30$, donc A est inversible et

$$A^{-1} = \begin{pmatrix} 1/10 & -1/5 \\ 2/15 & 1/15 \end{pmatrix}.$$

Le système étudié était $AX = B$, et nous pouvons maintenant multiplier par A^{-1} pour obtenir

$$A^{-1}(AX) = (A^{-1}A)X = \mathrm{Id}\,X = X = A^{-1}B = A^{-1}\begin{pmatrix} -9 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{-11}{10} \\ \frac{-17}{15} \end{pmatrix}.$$

En d'autres termes $x = {}^{-11}/_{10}$ et $y = {}^{-17}/_{15}$.

Les quaternions

Écrivons C pour l'ensemble des matrices 2×2, à coefficients dans $\mathbb{R}$, de la forme

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

On va écrire comme souvent I pour la matrice identité, et on pose

$$i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

(Les lettres minuscules sont rarement employées pour les matrices, mais vous allez voir pourquoi nous faisons une exception.) Avec ces notations, un élément de C s'écrit $a\mathrm{I} + bi$ de manière unique. De plus, faites le calcul, vous verrez que $i^2 = -\mathrm{I}$. Or l'élément I, nous l'avons signalé, se comporte « comme un 1 » dans les règles de calculs. Au total, l'ensemble C peut être identifié au corps $\mathbb{C}$ des nombres complexes, avec la règle $a\mathrm{I} + bi \leftrightarrow a + bi$, et cette identification est compatible avec les additions et les multiplications (en particulier, la somme et le produit de deux éléments de C sont encore dans C). Nous aurions très bien pu *définir* les nombres complexes comme ça, à l'aide des matrices.

On peut en fait aller plus loin. Soit maintenant $\mathbb{H}$ l'ensemble des matrices 4×4, à coefficients dans $\mathbb{R}$, de la forme

$$\begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}.$$

On écrit toujours I pour l'identité, puis on appelle i la matrice obtenue en faisant $b = 1$ et $a = c = d = 0$; on appelle j la matrice obtenue en faisant $c = 1$ et $a = b = d = 0$; et on appelle k la matrice obtenue en faisant $d = 1$ et $a = b = c = 0$. Ainsi tout $h \in \mathbb{H}$ s'écrit $h = a\mathrm{I} + bi + cj + dk$.

On vérifie que $i^2 = j^2 = k^2 = ijk = -\mathrm{I}$. Par suite, la somme et le produit de deux éléments de $\mathbb{H}$ sont encore dans $\mathbb{H}$ (par exemple en multipliant $ijk = -\mathrm{I}$ par k à droite, on obtient $ij = k$).

Notez que $ij = -ji$, donc $\mathbb{H}$ n'est pas commutatif. Par contre, on a la propriété suivante, très forte. Soit $h \in \mathbb{H}$ comme ci-dessus, et soit $\overline{h} = a\mathrm{I} - bi - cj - dk$. Alors $h\overline{h} = (a^2 + b^2 + c^2 + d^2)\mathrm{I}$. On en déduit que si $h \neq 0$, alors la matrice h est inversible, d'inverse $h^{-1} = ({}^{1}/_{\mathrm{N}})\overline{h}$, où $\mathrm{N} = a^2 + b^2 + c^2 + d^2$.

Voilà qui ne rentre pas dans la terminologie de la définition 2.17. On a les propriétés (a-b-c-d-e-f-g), c'est-à-dire que $\mathbb{H}$ est un anneau, mais on n'a pas (h) ; par contre on a (i). Ce type d'anneau est appelé un *corps gauche*, ou *corps non-commutatif*. (Dans certains livres, on dit « corps » pour corps gauche et « corps commutatif » pour ce que l'on appelle corps ici.)

Les éléments de $\mathbb{H}$ sont appelés les *quaternions*. On peut montrer, mais l'énoncé précis nous emmènerait déjà bien trop loin, qu'il n'y a essentiellement pas d'autre corps (gauche ou pas) « intéressant » contenant $\mathbb{R}$, à part $\mathbb{C}$ et $\mathbb{H}$. Évidemment il ne faut pas donner un sens trop littéral à cette dernière affirmation.

Définition 12.11 – Soit A une matrice. On dit que A est *échelonnée*, ou parfois *échelonnée en lignes*, lorsque les trois conditions suivantes sont satisfaites :

1. Dans chaque ligne de A, le premier coefficient non-nul (en partant de la gauche) est un 1. On l'appelle le *pivot* de la ligne.
2. À mesure que l'on descend dans les lignes, les pivots se décalent vers la droite.
3. Les lignes nulles de A sont situées en-dessous des lignes non-nulles.

De plus, on dit que A est *bien échelonnée* lorsqu'elle est échelonnée et que les pivots sont les seuls coefficients non-nuls dans leur colonne.

Exemple 12.12 – Les matrices suivantes sont échelonnées (les pivots sont encadrés) :

$$\begin{pmatrix} \boxed{1} & 4 & 0 & -3 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} \boxed{1} & 2 \\ 0 & \boxed{1} \end{pmatrix}.$$

La première est *bien* échelonnée, mais pas la deuxième (il faudrait que le 2 soit un 0).

Les matrices A_i ci-dessous *ne sont pas* échelonnées, car elles violent les règles 1, 2, 3 respectivement :

$$A_1 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 10 \end{pmatrix}.$$

Pourquoi prêter attention aux matrices (bien) échelonnées ? Tout simplement parce que les systèmes linéaires correspondants sont les plus simples possibles, en fait leurs solutions sont données sur un plateau.

Exemple 12.13 – Prenons la matrice suivante, qui est bien échelonnée :

$$A = \begin{pmatrix} \boxed{1} & 4 & 0 & -3 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Elle a quatre colonnes, elle peut donc décrire un système à quatre inconnues sur le modèle de l'exemple 12.6. En clair, posons

$$X = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \quad \text{et, par exemple,} \quad B = \begin{pmatrix} 4 \\ 2 \\ 0 \end{pmatrix},$$

alors le système AX = B s'écrit

$$\left\{ \begin{array}{rcrcrcrcl} \boxed{x} & + & 4y & & & - & 3t & = & 4 \\ & & & & \boxed{z} & & & = & 2 \\ & & & & & & 0 & = & 0 \end{array} \right. .$$

Ici les inconnues x et z sont encore appelées « pivots », on les a d'ailleurs encadrées. La matrice étant bien échelonnée, les solutions sont sous nos yeux :

- les inconnues qui ne sont pas des pivots vont servir de paramètres,
- les pivots vont être exprimés en fonction de ces paramètres.

Les paramètres sont donc y et t, et on a $z = 2$ et $x = 4-4y+3t$. En d'autres termes l'ensemble des solutions est

$$\left\{ \begin{pmatrix} 4-4y+3t \\ y \\ 2 \\ t \end{pmatrix} \text{ avec } y, t \in \mathbb{K} \right\} \subset \mathbb{K}^4.$$

Il est souvent plus lisible d'écrire cet ensemble sous la forme suivante :

$$\left\{ \begin{pmatrix} 4 \\ 0 \\ 2 \\ 0 \end{pmatrix} + y \begin{pmatrix} -4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 3 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ avec } y, t \in \mathbb{K} \right\}.$$

Cette méthode aurait fonctionné tout aussi bien pour n'importe quel second membre B, *sauf* dans les cas où c'est encore plus facile. Imaginons en effet que le dernier coefficient de B ne soit pas nul, disons

$$B = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

alors le système devient

$$\left\{ \begin{array}{rcrcrcrcl} \boxed{x} & + & 4y & & & - & 3t & = & 1 \\ & & & & \boxed{z} & & & = & 1 \\ & & & & & & 0 & = & 1 \end{array} \right. .$$

L'équation 0 = 1 étant impossible à satisfaire, le système n'a pas de solution du tout.

On vient de voir un système avec une infinité de solutions, puis un système sans solution. Sans changer la matrice A, il est clair que l'on tombe dans l'une ou l'autre de ces situations, selon le second membre B.

Mais il existe aussi des systèmes possédant une solution unique, par exemple avec

$$A' = \begin{pmatrix} 1 & -9 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

alors le système A′X = B s'écrit

$$\left\{ \begin{array}{rcrcl} x & - & 9y & = & b_1 \\ & & y & = & b_2 \end{array} \right. .$$

On a donc $y = b_2$ et $x = b_1 + 9y = b_1 + 9b_2$. Notez bien qu'ici la matrice A′ est échelonnée mais pas bien échelonnée (à cause du −9), et on a réussi à résoudre le système quand même. Ceci dit, on a dû faire une substitution supplémentaire (insérer la valeur de y dans x).

Il est clair que si un système AX = B avec A bien échelonnée n'a *pas* une infinité de solutions (donc pas de paramètres), alors il y a un pivot dans chaque colonne ; la matrice A ressemble donc à la matrice identité à laquelle on a rajouté des lignes de zéros. Ainsi, selon B, on a soit une solution unique (si les dernières équations sont toutes 0 = 0), soit aucune (s'il y en a une du type 0 = 1 comme ci-dessus).

En particulier, avec une matrice bien échelonnée, on n'obtient jamais de système ayant, disons, six solutions. On va voir que toutes ces observations faites sur les matrices échelonnées se généralisent aux matrices quelconques.

Puisqu'il est si facile de résoudre les systèmes correspondant aux matrices échelonnées, on aimerait pouvoir se ramener toujours à ce cas. C'est effectivement possible.

Définition 12.14 – Nous allons nous autoriser trois types d'opérations sur les lignes d'une matrice :

1. multiplier une ligne par un scalaire *non-nul*,
2. permuter deux lignes,
3. ajouter à une ligne un multiple d'une autre ligne.

Exemple 12.15 – Prenons la matrice

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 3 & 1 & 1 & -1 \\ 1 & 1 & -1 & 5 \end{pmatrix},$$

et faisons quelques opérations. Retranchons la première ligne à la dernière ligne (opération de type (3)) :

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 3 & 1 & 1 & -1 \\ 0 & 2 & 1 & 5 \end{pmatrix} \qquad L_3 \leftarrow L_3 - L_1$$

On a noté $L_3 \leftarrow L_3 - L_1$ pour indiquer l'opération effectuée ; d'autres abréviations sont possibles mais en tout cas il est bon d'indiquer au lecteur comment on a obtenu la nouvelle matrice. Poursuivons par une autre opération de type (3) :

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 0 & 4 & 7 & -1 \\ 0 & 2 & 1 & 5 \end{pmatrix} \qquad L_2 \leftarrow L_2 - 3L_1$$

Essayons une opération de type (1), puis une de type (2) :

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 0 & 4 & 7 & -1 \\ 0 & 1 & 1/2 & 5/2 \end{pmatrix} \qquad L_3 \leftarrow 1/2 L_3$$

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 0 & 1 & 1/2 & 5/2 \\ 0 & 4 & 7 & -1 \end{pmatrix} \qquad L_2 \leftrightarrow L_3$$

Essayons d'obtenir une matrice aussi simple que possible. Et d'abord, débarrassons-nous de ce 4 :

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 0 & 1 & 1/2 & 5/2 \\ 0 & 0 & 5 & -11 \end{pmatrix} \qquad L_3 \leftarrow L_3 - 4L_2$$

$$\begin{pmatrix} \boxed{1} & -1 & -2 & 0 \\ 0 & \boxed{1} & 1/2 & 5/2 \\ 0 & 0 & \boxed{1} & -11/5 \end{pmatrix} \qquad L_3 \leftarrow \frac{1}{5} L_3$$

C'est déjà une matrice échelonnée. On peut même poursuivre notre effort et obtenir une matrice bien échelonnée :

$$\begin{pmatrix} 1 & -1 & -2 & 0 \\ 0 & 1 & 0 & 18/5 \\ 0 & 0 & 1 & -11/5 \end{pmatrix} \qquad L_2 \leftarrow L_2 - \frac{1}{2} L_3$$

$$\begin{pmatrix} 1 & -1 & 0 & -22/5 \\ 0 & 1 & 0 & 18/5 \\ 0 & 0 & 1 & -11/5 \end{pmatrix} \qquad L_1 \leftarrow L_1 + 2L_3$$

et enfin

$$\begin{pmatrix} \boxed{1} & 0 & 0 & -4/5 \\ 0 & \boxed{1} & 0 & 18/5 \\ 0 & 0 & \boxed{1} & -11/5 \end{pmatrix} \qquad L_1 \leftarrow L_1 + L_2$$

Théorème 12.16 – *En faisant des opérations sur les lignes d'une matrice* A*, on peut toujours obtenir une matrice bien échelonnée, et une seule.*

On notera E_A la matrice bien échelonnée associée à A.

L'unicité de la matrice E_A sera démontrée dans la deuxième moitié de ce chapitre. Pour l'existence, il suffit de procéder comme dans l'exemple 12.15. On peut garder en tête les lignes directrices suivantes :

- On choisit une ligne avec un coefficient non-nul dans la première colonne, on met cette ligne en première position, puis avec des opérations de type (3) on fait apparaître des 0 dans la première colonne sur les autres lignes. On divise la première ligne par son premier coefficient pour que celui-ci devienne un 1.
- On oublie la première colonne et la première ligne complètement, et on continue avec le reste de la matrice. On obtient alors une matrice échelonnée.
- Pour obtenir une matrice *bien* échelonnée, on commence par le dernier pivot. Avec des opérations de type (3), on fait apparaître des 0 au-dessus de ce pivot. Puis on recommence avec les pivots précédents.

Ou alors, relisez l'exemple 12.15 soigneusement.

exercice : mettez sous forme échelonnée toutes les matrices des exercices précédents

Toutes ces considérations n'auraient pas un grand intérêt vis-à-vis des systèmes si l'on n'avait pas le résultat suivant, qui affirme que les opérations sur les lignes ne changent pas les solutions :

Proposition 12.17 – *Considérons un système de la forme* AX = B*, où* X *et* B *sont des colonnes, et* X *contient les inconnues.*

Soit A′ *obtenue à partir de* A *en faisant des opérations sur les lignes, et soit* B′ *obtenue en faisant les mêmes opérations sur* B*. Alors le système* A′X = B′ *a les mêmes solutions que* AX = B*.*

Là encore, la démonstration sera donnée plus loin dans ce chapitre (elle n'est pas difficile).

Exemple 12.18 – Considérons le système

$$\left\{ \begin{array}{rcrcrcr} x & - & y & - & 2z & = & 0 \\ 3x & + & y & + & z & = & -1 \\ x & + & y & - & z & = & 5 \end{array} \right. .$$

Il est de la forme AX = B en posant

$$A = \begin{pmatrix} 1 & -1 & -2 \\ 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}, \quad X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ -1 \\ 5 \end{pmatrix}.$$

D'après la proposition, on ne change pas les solutions en faisant des opérations sur A et B en même temps. Il s'agit de faire les *mêmes* opérations sur ces deux matrices, et en pratique de nombreux étudiants préfèrent ajouter B comme colonne à A, de sorte que l'on fait des opérations sur la matrice

$$\left(\begin{array}{rrr|r} 1 & -1 & -2 & 0 \\ 3 & 1 & 1 & -1 \\ 1 & 1 & -1 & 5 \end{array}\right).$$

(La ligne verticale est juste là pour rappeler d'où vient la dernière colonne.) C'est la matrice sur laquelle nous avons travaillé dans l'exemple 12.15. Nous avons vu qu'après quelques opérations, on arrive à

$$\left(\begin{array}{ccc|c} \boxed{1} & 0 & 0 & -4/5 \\ 0 & \boxed{1} & 0 & 18/5 \\ 0 & 0 & \boxed{1} & -11/5 \end{array}\right).$$

En faisant la traduction inverse, on retrouve un système, qui est particulièrement simple :

$$\left\{ \begin{array}{ccccc} x & & & = & -4/5 \\ & y & & = & 18/5 \\ & & z & = & -11/5 \end{array} \right.$$

Ces trois valeurs de x, y et z constituent l'unique solution du système initial AX = B, comme vous pouvez le vérifier.

voir les exercices 1171, 1173

Nous allons donner une méthode pour calculer l'inverse d'une matrice quand elle existe (ou démontrer que l'inverse n'existe pas lorsque c'est le cas). Nous ne dirons rien de la démonstration ici : c'est encore un résultat établi plus loin dans ce chapitre.

abus fréquent : « inverse » au féminin comme raccourci de « la matrice inverse »

La méthode va paraître un peu magique. Donnons simplement l'indication suivante : nous venons d'expliquer comment résoudre les systèmes linéaires en faisant des opérations sur les lignes, alors que dans l'exemple 12.10 nous indiquions comment résoudre un système si l'on connaît l'inverse de sa matrice. C'est en comparant ces deux méthodes, qui doivent bien donner le même résultat, qu'on en déduit la méthode de calcul. Voici le principe.

PROPOSITION 12.19 – *Soit* $A \in M_n(\mathbb{K})$. *Alors :*

1. *A est inversible si et seulement si sa matrice bien échelonnée* E_A *est l'identité.*
2. *Étant donnée une suite d'opérations sur les lignes qui transforment* A *en* $E_A = \mathrm{Id}$, *on obtient* A^{-1} *en faisant les mêmes opérations (dans le même ordre) sur la matrice identité.*

EXEMPLE 12.20 – Voyons comment mettre ceci en pratique. Admettons que l'on s'intéresse à l'inverse de la matrice

$$A = \begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix}.$$

On va faire des opérations sur les lignes de A pour trouver sa forme bien échelonnée, et chaque opération est faite en parallèle sur la matrice identité. On commence donc par présenter les matrices côte à côte :

$$\begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

On commence :

$$\begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 3 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad L_3 \leftarrow L_3 + L_1$$

$$\begin{pmatrix} -1 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 5 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix} \qquad L_3 \leftarrow L_3 + 3L_2$$

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1/5 & 3/5 & 1/5 \end{pmatrix} \qquad \begin{matrix} L_1 \leftarrow -L_1 \\ L_2 \leftarrow -L_2 \\ L_3 \leftarrow \frac{1}{5}L_3 \end{matrix}$$

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -6/5 & -3/5 & -1/5 \\ 1/5 & -2/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \end{pmatrix} \qquad \begin{matrix} L_2 \leftarrow L_2 + L_3 \\ L_1 \leftarrow L_1 - L_3 \end{matrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & -1 & 0 \\ 1/5 & -2/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \end{pmatrix} \qquad L_1 \leftarrow L_1 + L_2$$

On a obtenu une matrice bien échelonnée, c'est donc E_A. Ici on a $E_A = \mathrm{Id}$, donc A est inversible d'après la proposition. Cette même proposition affirme aussi que la matrice A^{-1} est la dernière matrice écrite à droite, c'est-à-dire

$$A^{-1} = \begin{pmatrix} -1 & -1 & 0 \\ 1/5 & -2/5 & 1/5 \\ 1/5 & 3/5 & 1/5 \end{pmatrix}.$$

Prenons un autre exemple, disons

$$B = \begin{pmatrix} 1 & 2 & -1 \\ -6 & 16 & -2 \\ -11 & 6 & 3 \end{pmatrix}.$$

C'est reparti :

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 28 & -8 \\ 0 & 28 & -8 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ 11 & 0 & 1 \end{pmatrix} \qquad \begin{matrix} L_2 \leftarrow L_2 + 6L_1 \\ L_3 \leftarrow L_3 + 11L_1 \end{matrix}$$

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 28 & -8 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ 5 & -1 & 1 \end{pmatrix} \qquad L_3 \leftarrow L_3 - L_2$$

Arrêtons-nous : on vient d'obtenir une ligne de zéros. En effet, si l'on poursuivait le calcul on obtiendrait une matrice bien échelonnée qui elle-même aurait une ligne de zéros, donc $E_B \neq \mathrm{Id}$. D'après la proposition, la matrice B n'est pas inversible. Les calculs que l'on a faits sur la matrice identité n'auront servi à rien : c'est l'un des petits défauts de la méthode.

voir les exercices 2750, 2775, 2773

– Fin de la première lecture –

L'observation suivante est riche de conséquences : faire une opération sur les lignes revient à multiplier à gauche par une matrice inversible. Plus précisément :

PROPOSITION 12.21 – *Soient* A *une matrice, et* A′ *obtenue en faisant une opération sur les lignes de* A. *Alors il existe une matrice inversible* P *telle que* $A' = PA$.

De plus, pour chaque type d'opération on peut trouver une matrice P *unique qui convient pour toutes les matrices* A *à la fois.*

Démonstration. Cherchons P qui convient pour toutes les matrices A. On n'a pas beaucoup de choix, puisqu'en faisant $A = \mathrm{Id}$, on a $A' = P\,\mathrm{Id} = P$: en d'autres termes, la matrice P elle-même doit être obtenue en faisant l'opération en question sur les lignes de la matrice identité.

Par exemple, pour multiplier la première ligne par $\lambda \neq 0$, on doit prendre

$$P = M_\lambda = \begin{pmatrix} \lambda & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

Or on vérifie que pour toute matrice A, le produit $M_\lambda A$ est effectivement obtenu en multipliant la première ligne de A par λ. De plus M_λ est bien inversible, d'inverse $M_{\lambda^{-1}}$. Donc la proposition est vraie pour cette opération. Pour multiplier une autre ligne, déplacer λ le long de la diagonale.

Pour permuter la première ligne et la deuxième, on doit prendre

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix}.$$

On vérifie que cette matrice convient effectivement. Elle est inversible, et même égale à on inverse.

Pour ajouter λ fois la deuxième ligne à la première, prendre

$$P = C_\lambda^{12} = \begin{pmatrix} 1 & \lambda & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

Cette matrice convient, et son inverse est $C_{-\lambda}^{12}$.

Les autres cas sont similaires. □

COROLLAIRE 12.22 – *Pour chaque matrice* A, *il existe une matrice* P *inversible telle que* $E_A = PA$.

Attention, cette matrice P dépend de A fortement !

Démonstration. On peut obtenir E_A en faisant des opérations sur les lignes de A ; disons que cela nécessite k étapes. Si la première opération correspond à la matrice P_1, alors après une étape on travaille avec P_1A. Si la deuxième opération est donnée par P_2, on se retrouve avec P_2P_1A. Après k opérations, on a $P_kP_{k-1}\cdots P_1A = E_A$. La matrice $P = P_kP_{k-1}\cdots P_1$ est inversible et son inverse est $P_1^{-1}P_2^{-1}\cdots P_k^{-1}$. □

On peut maintenant démontrer très facilement la proposition 12.17 :

Démonstration de la proposition 12.17 . Pour toute matrice inversible P, on a

$$AX = B \iff PAX = PB,$$

puisqu'on passe de l'égalité de gauche à celle de droite en multipliant par P, et dans l'autre sens en multipliant par P^{-1}. Or faire des opérations sur les lignes revient bien à multiplier par une matrice inversible. □

Justification de la méthode de calcul de l'inverse

Commençons par un petit lemme utile.

Lemme 12.23 – *Soient* A *et* B *deux matrices carrées telles que*

$$AB = \mathrm{Id}\,.$$

Alors A *et* B *sont toutes les deux inversibles, et inverses l'une de l'autre.*

Rappelez-vous que dans la définition 12.8, on donnait deux conditions à vérifier : $AB = \mathrm{Id}$ et également $BA = \mathrm{Id}$. Donc ce résultat affirme qu'une seule de ces conditions entraîne l'autre.

Démonstration. Soit P inversible telle que $PA = E_A$. On a donc

$$PAB = E_A B = P\,\mathrm{Id} = P\,.$$

En particulier la matrice $E_A B = P$ est inversible. On en conclut qu'il ne peut pas y avoir de lignes nulles dans E_A, sinon il en serait de même dans $E_A B$, et cette matrice ne pourrait pas être inversible.

Puisque E_A est bien échelonnée sans lignes nulles, et carrée, on doit avoir $E_A = \mathrm{Id}$. Ainsi $E_A B = \mathrm{Id}\,B = B = P$, et B est inversible. En multipliant $AB = \mathrm{Id}$ par B^{-1} à droite, on obtient $A = B^{-1}$, donc A est inversible également, et c'est l'inverse de B. □

Démonstration de la propositon 12.19. Si A est inversible, on a $E_A = \mathrm{Id}$ comme on vient de le voir dans la démonstration du lemme. Réciproquement si $E_A = \mathrm{Id}$, alors on prend P inversible telle que $PA = E_A = \mathrm{Id}$. D'après le lemme A est inversible et son inverse est $A^{-1} = P$. Ceci établit déjà le (1) de la proposition 12.19.

Puisque $P = A^{-1}$, on a $P\,\mathrm{Id} = A^{-1}$. Or multiplier à gauche par P revient à faire des opération sur les lignes, et on constate bien que A^{-1} s'obtient en faisant sur la matrice identité les mêmes opérations que l'on a faites sur A. C'est ce que dit le (2) de la proposition. □

Proposition 12.24 – *Soient* E_1 *et* E_2 *des matrices bien échelonnées de même taille. On suppose qu'il existe une matrice* P *inversible telle que* $E_2 = PE_1$. *Alors* $E_1 = E_2$.

Démonstration. Si $E_1 = 0$, alors $E_2 = 0 = E_1$, donc on envisage la situation où $E_1 \neq 0$. Supposons que E_1 commence par k colonnes nulles (avec éventuellement $k = 0$). La $k+1$-ème colonne est donc

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

puisque E_1 est bien échelonnée. En conséquence la matrice PE_1 commence également par k colonnes de 0, et sa $(k+1)$-ème colonne est la première colonne de P. Mais $PE_1 = E_2$ est bien échelonnée, donc cette colonne de P est elle aussi de la forme

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

On voit donc déjà que E_1 et E_2 sont identiques dans les $k+1$ premières colonnes. Plus précisément on écrit

$$E_i = \left(\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & & & \\ 0 & \cdots & 0 & 0 & & F_i & \\ \vdots & & \vdots & \vdots & & & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right)$$

pour $i = 1$ ou 2. On observe que la matrice F_i est bien échelonnée. De même on note

$$P = \left(\begin{array}{c|cccc} 1 & a_2 & a_3 & \cdots & a_n \\ 0 & & & & \\ 0 & & P' & & \\ \vdots & & & & \\ 0 & & & & \end{array}\right).$$

Cette matrice P′ est inversible ; en fait, P étant une matrice $n \times n$, alors l'inverse de P′ est le bloc $(n-1) \times (n-1)$ en bas à droite de P^{-1}.

À partir de l'égalité $E_2 = PE_1$, on tire facilement $F_2 = P'F_1$. Faisons une récurrence sur le nombre de lignes des matrices (le cas des matrices-lignes vient d'être traité en silence, puisqu'on a établi que $P = [1]$ dans cette situation). Comme F_i est strictement plus petite que E_i, on peut supposer que l'on connaît la proposition dans ce cas, et donc que $F_1 = F_2$. Reste à montrer que E_1 et E_2 ont la même première ligne, sachant que la première ligne de $E_2 = PE_1$ est

$$L_1 + a_2 L_2 + \cdots + a_n L_n \tag{*}$$

où L_i est la i-ème ligne de E_1.

Si E_1 a un pivot dans la ligne $i > 1$, alors E_2 aussi puisque $F_1 = F_2$. Un pivot étant seul dans sa colonne, on constate que, dans la même colonne de PE_1, on trouve a_i sur la première ligne. Puisque $PE_1 = E_2$ est bien échelonnée avec un pivot dans cette colonne, on doit avoir $a_i = 0$ dans ce cas, donc $a_i L_i = 0$.

Si par contre la ligne L_i de E_1 n'a pas de pivot, c'est qu'elle est nulle. On a encore $a_i L_i = 0$.

Finalement, l'expression (*) montre clairement que la première ligne de $E_2 = PE_1$ est L_1, qui est aussi la première ligne de E_1, comme on le souhaitait. □

Démonstration du théorème 12.16. Si E_1 et E_2 sont deux matrices bien échelonnées obtenues à partir de A, alors il existe des matrices inversibles P_1 et P_2 telles que $P_1 A = E_1$ et $P_2 A = E_2$ (proposition 12.21). Ainsi $E_2 = P_2 P_1^{-1} E_1$, donc la proposition précédente montre que $E_2 = E_1$. La matrice échelonnée associée à A est bien unique. □

Chapitre 13

Déterminants

On rappelle que la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$.

L'objectif de ce chapitre est de montrer une généralisation de la proposition 12.9. Plus précisément, on cherche à associer à toute matrice carrée un nombre, son *déterminant*, qui soit facilement calculable et qui permette de décider si la matrice est inversible ou non. On aimerait aussi avoir une formule pour l'inverse (bien que l'on sache déjà calculer les inverses efficacement).

C'est dans la deuxième partie de ce chapitre que nous montrerons le théorème suivant :

Théorème 13.1 – *Il existe une unique fonction*

$$\begin{array}{rccc} \det\colon & M_n(\mathbb{K}) & \longrightarrow & \mathbb{K} \\ & A & \mapsto & \det(A) \end{array}$$

ayant les propriétés suivantes :

1. *Si* A_1 *est obtenue à partir de* A *en multipliant une ligne par* $\lambda \in \mathbb{K}$*, alors* $\det(A_1) = \lambda \det(A)$.
2. *Si* A_2 *est obtenue à partir de* A *en échangeant deux lignes, alors* $\det(A_2) = -\det(A)$.
3. *Si* A_3 *est obtenue à partir de* A *en ajoutant à une ligne un multiple d'une autre ligne, alors* $\det(A_3) = \det(A)$.
4. $\det(\mathrm{Id}) = 1$.

De plus, pour toute matrice A*, on a* $\det(A) = \det({}^tA)$*. Par suite, on peut remplacer « ligne » par « colonne » dans ce qui précède.*

Enfin, si $\phi\colon M_n(\mathbb{K}) \to \mathbb{K}$ *est une fonction ayant les propriétés (1), (2) et (3) ci-dessus (mais pas forcément (4)), alors il existe un nombre* $\alpha \in \mathbb{K}$ *tel que* $\phi(A) = \alpha \det(A)$.

C'est tout ce que nous avons besoin de savoir sur cette fonction, et le fait de rejeter la définition du déterminant à plus tard ne va nous empêcher ni de les calculer, ni de montrer leur utilité.

Commençons par quelques calculs. Ensuite nous montrerons que les déterminants ont bien quelque chose à voir avec les inverses.

Exemple 13.2 – Pour les matrices 2×2, la fonction que l'on a déjà vue

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc$$

vérifie les quatre conditions : vérifiez-le ! Donc par unicité, c'est bien la fonction déterminant :

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc\,.$$

Exemple 13.3 – Prenons

$$M = \begin{pmatrix} 2 & 2 & 2 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix},$$

et calculons det(M). On va utiliser les propriétés ci-dessus pour se ramener à une matrice échelonnée, comme nous savons le faire.

D'abord une notation : on va écrire

$$\begin{vmatrix} 2 & 2 & 2 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = \det \begin{pmatrix} 2 & 2 & 2 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

La propriété (1) est souvent interprétée « à l'envers » par les étudiants, au début. Il s'agit bien de la chose suivante : la matrice M est obtenue en multipliant par 2 la première ligne de

$$N = \begin{pmatrix} 1 & 1 & 1 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

On a donc $\det(M) = 2\det(N)$. La règle est simple : on « sort » le 2 du déterminant. On va écrire les choses comme ceci (on continue avec les autres lignes) :

$$\begin{vmatrix} 2 & 2 & 2 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = 2 \begin{vmatrix} 1 & 1 & 1 \\ -3 & 3 & -3 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{vmatrix}$$

$$= 2 \times 3 \begin{vmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = 2 \times 3 \times \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \end{vmatrix}.$$

Maintenant faisons une opération de type (3) :

$$\begin{vmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & -1 \end{vmatrix} \qquad L_2 \leftarrow L_2 + L_1\,.$$

Ces opérations ne changent pas le déterminant. On continue :

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 1 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{vmatrix} \qquad L_3 \leftarrow L_3 - L_1\,.$$

Avec l'habitude, vous verrez immédiatement que ce dernier déterminant vaut −4. Pourquoi ? Voyons :

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{vmatrix} = 2 \times (-2) \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$= -4 \begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -4 \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

Pour les deux dernières égalités, on a fait successivement les opérations $L_1 \leftarrow L_1 - L_3$ puis $L_1 \leftarrow L_1 - L_2$.

Le déterminant de la matrice identité vaut 1 d'après le théorème, donc

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{vmatrix} = -4,$$

comme annoncé. Quant au déterminant de M, il vaut finalement

$$\det(M) = 2 \times 3 \times \frac{1}{2} \times (-4) = -12\,.$$

Exemple 13.4 – Avec des opérations sur les lignes, on peut toujours se ramener au calcul du déterminant d'une matrice bien échelonnée. Que dire s'il ne s'agit pas de l'identité ? Que vaut par exemple det(A) lorsque

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad ?$$

C'est bien simple : en multipliant la troisième ligne de A par 0, on obtient encore la matrice A. Donc $\det(A) = 0 \times \det(A) = 0$. *Le déterminant d'une matrice ayant une ligne nulle est nul*. Et pareil avec les colonnes.

voir les exercices 1134, 1148, 1136

Le résultat suivant est alors facile :

Proposition 13.5 – *Une matrice carrée est inversible si et seulement si son déterminant est non-nul.*

Démonstration. Si A′ est obtenue à partir de A en faisant une opération sur les lignes, on a $\det(A') \neq 0 \Leftrightarrow \det(A) \neq 0$ (on se rappellera que les opérations « autorisées » pour échelonner une matrice, données dans la définition 12.14, comprennent la multiplication d'une ligne par un scalaire *non-nul*). Si E_A désigne comme d'habitude la matrice bien échelonnée associée à A, on constate que $\det(A) \neq 0$ si et seulement si $\det(E_A) \neq 0$.

De plus, pour une matrice bien échelonnée telle que E_A, on constate que son déterminant vaut 0 s'il y a une ligne nulle, et 1 sinon, puisque le seul cas dans lequel il n'y a pas de ligne nulle dans E_A se produit pour $E_A = \mathrm{Id}$.

Pour finir on sait que A est inversible si et seulement si $E_A = \mathrm{Id}$ (proposition 12.19), donc si et seulement si $\det(E_A) = 1$, ce qui se produit si et seulement si $\det(A) \neq 0$. □

Par exemple, la matrice M de l'exemple 13.3 est inversible puisque son déterminant vaut $-12 \neq 0$. Pour montrer ça, nous avons en fait échelonné la matrice et trouvé au passage que $E_M = \mathrm{Id}$, ce qui établit directement que M est inversible. On se demande si l'on y gagne. Il est vrai qu'avec l'habitude, on calcule les déterminants assez rapidement en multipliant les astuces et les raccourcis lorsqu'on reconnaît certaines matrices – ça sera particulièrement vrai lorsque nous donnerons la deuxième méthode de calcul, qui se prête au calcul mental rapide.

Il ne faut pas trop prendre ça au sérieux, cependant : calculatoirement, les déterminants ne sont pas un outil révolutionnaire. Pour une matrice un peu grosse (quelques milliers de lignes), dont on veut savoir si elle est inversible à l'aide d'un ordinateur, la meilleure méthode consiste à faire des opérations sur les lignes pour obtenir la matrice échelonnée, et les « raccourcis » que les déterminants permettent de prendre n'ont pas une influence sensible sur la durée du calcul.

Les déterminants ont pourtant de nombreuses applications. Donnons-en une simple.

Exemple 13.6 – Soit $t \in \mathbb{K}$. Quand est-ce que la matrice

$$\begin{pmatrix} 2 & t \\ -1 & 9 \end{pmatrix}$$

est inversible ? Son déterminant vaut $18 + t$, donc la condition est précisément $t \neq -18$. On peut retrouver cette condition sans passer par le déterminant, mais c'est moins facile à suivre.

Pour un exemple plus compliqué, soit θ un paramètre réel. Quand est-ce que la matrice

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

est inversible ? En prenant le déterminant, on obtient $\cos(\theta)^2 + \sin(\theta)^2 = 1$, donc la matrice est toujours inversible. En faisant des opérations sur les lignes pour trouver la matrice échelonnée, on se retrouve dans une discussion très pénible.

La proposition suivante rend souvent service.

Proposition 13.7 – *Soient* A *et* B *deux matrices carrées de même taille. Alors*

$$\det(AB) = \det(A)\det(B)\,.$$

De plus si A *est inversible alors*

$$\det(A^{-1}) = \frac{1}{\det(A)}\,.$$

Démonstration. La matrice B étant fixée, considérons la fonction ϕ définie par $\phi(A) = \det(AB)$. On voit tout de suite qu'elle vérifie les propriétés (1), (2) et (3) du théorème 13.1. (Noter que faire des opérations sur les lignes de AB revient à faire des opérations sur les lignes de A, puis multiplier par B.) D'après le théorème, il existe une constante α telle que $\phi(A) = \alpha \det(A)$.

En prenant A = Id, on observe $\phi(\mathrm{Id}) = \det(\mathrm{Id}\,B) = \det(B) = \alpha\det(\mathrm{Id}) = \alpha$, c'est-à-dire $\alpha = \det(B)$. Ceci montre que $\det(AB) = \det(A)\det(B)$.

voir l'exercice 1144

Si A^{-1} existe, on calcule $\det(AA^{-1}) = \det(A)\det(A^{-1}) = \det(\mathrm{Id}) = 1$. □

Définition 13.8 – Soit A une matrice $n \times n$. On appelle *mineur en* i, j, et on note Δ_{ij}, le déterminant de la matrice $(n-1)\times(n-1)$ obtenue en supprimant la ligne i et la colonne j de A.

Exemple 13.9 – Soit

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Alors

$$\Delta_{11} = \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix}, \ \Delta_{22} = \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix}, \ \Delta_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

Définition 13.10 – Notons $A = (a_{ij})$. Le *développement de* $\det(A)$ *par la ligne* i est

$$\det_i(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \Delta_{ij}.$$

Le *développement de* $\det(A)$ *par la colonne* j est

$$\det^j(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \Delta_{ij}.$$

(Cette fois c'est l'indice i qui varie dans la somme.)

Exemple 13.11 – Prenons

$$A = \begin{pmatrix} 0 & -1 & -1 \\ -2 & 0 & 5 \\ -3 & 0 & -1 \end{pmatrix}.$$

Pour gérer le signe $(-1)^{i+j}$, le plus simple est de former un tableau :

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

La règle est simple : on commence par + en haut à gauche, et on alterne de sorte qu'il n'y a jamais deux signes identiques côte à côte.

Pour développer par rapport à la première ligne, disons, on prend les mineurs et les coefficients dans la matrice, les signes dans le tableau, et on ajoute :

$$\det_1(A) = +0 \times \begin{vmatrix} 0 & 5 \\ 0 & -1 \end{vmatrix} - (-1) \times \begin{vmatrix} -2 & 5 \\ -3 & -1 \end{vmatrix} + (-1) \times \begin{vmatrix} -2 & 0 \\ -3 & 0 \end{vmatrix}$$

$$= 0 + 17 + 0 = 17.$$

Par rapport à la troisième colonne, on a

$$\det^3(A) = +(-1) \times \begin{vmatrix} -2 & 0 \\ -3 & 0 \end{vmatrix} - 5 \times \begin{vmatrix} 0 & -1 \\ -3 & 0 \end{vmatrix} + (-1) \times \begin{vmatrix} 0 & -1 \\ -2 & 0 \end{vmatrix}$$

$$= 0 - 5 \times (-3) - (-2) = 17.$$

Ce n'est pas un hasard si on trouve le même résultat.

Proposition 13.12 – *Pour tout i et tout j, on a*

$$\det_i(A) = \det^j(A) = \det(A).$$

En d'autre termes, le développement donne une formule pour calculer le déterminant, en fonction de déterminants plus petits.

Démonstration. Pour l'instant nous allons nous contenter d'une esquisse de démonstration. Dans la suite du chapitre un argument complet sera donné.

L'idée simple est que $A \mapsto \det_i(A)$, vue comme une fonction $M_n(\mathbb{K}) \to \mathbb{K}$, vérifie les propriétés (1), (2), (3) et (4) du théorème 13.1. D'après ce théorème, une telle fonction est unique, donc $\det_i(A) = \det(A)$. On procède de même pour $\det^j(A)$.

Il faut donc vérifier soigneusement ces quatre propriétés. Ce n'est pas difficile (c'est même un bon exercice), mais c'est un peu long. □

Corollaire 13.13 – *Le déterminant d'une matrice peut s'exprimer en fonction des coefficients par des opérations algébriques simples (additions et multiplications). En particulier, si une matrice $A \in M_n(\mathbb{C})$ a ses coefficients dans $\mathbb{R}$, $\mathbb{Q}$ ou même $\mathbb{Z}$, alors son déterminant est un élément de $\mathbb{R}$, $\mathbb{Q}$ ou $\mathbb{Z}$ respectivement.*

Démonstration. C'est vrai pour les matrices 2×2 de par la formule $ad - bc$, et grâce à la proposition on peut faire une récurrence. Le corollaire sera par ailleurs évident lorsqu'on aura donné une définition du déterminant, plus loin dans ce chapitre. □

On a donc une nouvelle méthode, complètement différente, pour calculer les déterminants. Est-elle meilleure ? En théorie la réponse est cinglante : non, elle est même bien pire. Pour s'en convaincre, comptons le nombre de multiplications nécessaires pour calculer un déterminant de taille $n \times n$. En échelonnant la matrice, on peut montrer qu'il faut quelque chose de l'ordre de n^2 multiplications pour avoir le déterminant. En développant par une ligne, puis en développant les déterminants plus petits, etc., jusqu'à en arriver à des déterminants 2×2, on fait plus de $n!$ opérations.

Prenons $n = 50$. À l'aide d'un ordinateur qui effectuerait un milliard de milliards d'opérations par seconde (10^{18}), il faudrait plus de 10^{38} ans pour compléter le calcul par cette méthode. L'univers selon les dernières estimations existe depuis 15×10^9 ans. En échelonnant la matrice, les 2500 multiplications nécessaires sont faites presque instantanément.

En pratique, pour des matrices de très petite taille ($n \leq 6$) comme vous en rencontrerez dans les exercices, et sans ordinateur, on combine souvent différentes méthodes : opérations sur les lignes, sur les colonnes, et quelques développements. Voyons un exemple.

Exemple 13.14 – Prenons

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 2 & -1 & 9 & 0 \\ 2 & 7 & 4 & -5 \\ 2 & 0 & -1 & -1 \end{pmatrix}.$$

Pour calculer $\det(A)$, on peut commencer par retrancher la deuxième colonne à la dernière, ce qui ne change pas le déterminant. On obtient

$$\det(A) = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 2 & -1 & 9 & 1 \\ 2 & 7 & 4 & -12 \\ 2 & 0 & -1 & -1 \end{vmatrix}.$$

Maintenant on développe par la première ligne, évidemment, puisqu'elle a tellement de zéros :

$$\det(A) = -\begin{vmatrix} 2 & 9 & 1 \\ 2 & 4 & -12 \\ 2 & -1 & -1 \end{vmatrix},$$

en n'oubliant pas le signe – qui vient du tableau de signes correspondant.

On retranche la première ligne aux suivantes :

$$\det(A) = -\begin{vmatrix} 2 & 9 & 1 \\ 0 & -5 & -13 \\ 0 & -10 & -2 \end{vmatrix},$$

puis on effectue $L_3 \leftarrow L_3 - 2L_2$:

$$\det(A) = -\begin{vmatrix} 2 & 9 & 1 \\ 0 & -5 & -13 \\ 0 & 0 & 24 \end{vmatrix}.$$

Nous sommes déjà habitués aux déterminants de cette forme « triangulaire ». On peut mentalement faire le développement suivant par la première colonne :

$$\begin{vmatrix} 2 & 9 & 1 \\ 0 & -5 & -13 \\ 0 & 0 & 24 \end{vmatrix} = 2 \begin{vmatrix} -5 & -13 \\ 0 & 24 \end{vmatrix} = 2 \times (-5) \times 24 = -240.$$

(D'une manière générale, le déterminant d'une matrice triangulaire est le produit des coefficients sur la diagonale.) Finalement $\det(A) = 240$.

reprendre les déterminants des exercices précédents par cette nouvelle méthode

Nous allons maintenant voir une formule pour calculer l'inverse d'une matrice, qui est une généralisation de celle donnée dans la proposition 12.9.

Soit $A = (a_{ij})$ une matrice. On pose

$$\tilde{A} = ((-1)^{i+j}\Delta_{ij})_{ij},$$

et on l'appelle la *comatrice* de A. (On rappelle que le « mineur » Δ_{ij} a été introduit dans la définition 13.8.)

Proposition 13.15 – *On a*

$$A\,{}^t\tilde{A} = \det(A)\,\mathrm{Id}.$$

Démonstration. Sur la ligne i, dans la colonne j du produit $A\,{}^t\tilde{A}$, on a par définition le nombre

$$c_{ij} = \sum_{k=1}^{n} a_{ik}(-1)^{k+j}\Delta_{jk}.$$

Lorsque $i = j$, on reconnaît la formule pour le développement de det(A) par la ligne i, donc $c_{ii} = \det(A)$. Lorsque $i \neq j$ par contre, on obtient le développement de det(A′) par la ligne j, où A′ est la matrice obtenue à partir de A en recopiant la ligne i dans la ligne j ; comme A′ se retrouve avec deux lignes identiques, son déterminant est nul, donc $c_{ij} = 0$ pour $i \neq j$.

Finalement $A\,{}^t\tilde{A}$ n'a de coefficients non-nuls que sur la diagonale, où l'on trouve det(A), comme annoncé. □

Corollaire 13.16 (Formules de Cramer) – *Soit* A *une matrice inversible. Alors*

$$A^{-1} = \frac{1}{\det(A)}\,{}^t\tilde{A}.$$

Démonstration. D'après la proposition, on a

$$A\left(\frac{1}{\det(A)}\,{}^t\tilde{A}\right) = \mathrm{Id},$$

ce qui donne le résultat (voir lemme 12.23). □

Exemple 13.17 – Prenons

$$A = \begin{pmatrix} 1 & -3 & -1 \\ 1 & -4 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Pour calculer la comatrice, on commence par le coefficient en haut à gauche, qui doit être

$$\begin{vmatrix} -4 & 0 \\ 1 & 2 \end{vmatrix} = -8.$$

Le coefficient sur la ligne 1, dans la colonne 2, est

$$-\begin{vmatrix} 1 & 0 \\ 0 & 2 \end{vmatrix} = -2.$$

(On n'oublie pas le signe.) Ainsi de suite, on finit par obtenir

$$\tilde{A} = \begin{pmatrix} -8 & -2 & 1 \\ 5 & 2 & -1 \\ -4 & -1 & -1 \end{pmatrix}.$$

On calcule $\det(A) = -3$. Finalement

$$A^{-1} = -\frac{1}{3}\begin{pmatrix} -8 & 5 & -4 \\ -2 & 2 & -1 \\ 1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 8/3 & -5/3 & 4/3 \\ 2/3 & -2/3 & 1/3 \\ -1/3 & 1/3 & 1/3 \end{pmatrix}.$$

Cette méthode de calcul de l'inverse est déraisonnablement populaire auprès des étudiants. Pourtant, elle est beaucoup moins efficace que la méthode de la proposition 12.19 (on peut de nouveau compter le nombre d'opérations effectuées, et il est largement supérieur avec les formules de Cramer). C'est pourquoi nous n'insisterons pas sur les exemples.

Par contre les formules de Cramer ont des conséquences plus théoriques, par exemple :

Lemme 13.18 – *Soit* A *une matrice à coefficients dans* $\mathbb{Z}$. *Alors* A *possède une inverse à coefficients dans* $\mathbb{Z}$ *si et seulement si* $\det(A) = \pm 1$.

Démonstration. Si $\det(A) = \pm 1$, la formule pour A^{-1} montre bien, en gardant le corollaire 13.13 à l'esprit, que ses coefficients sont des nombres entiers.

Réciproquement, si A^{-1} est à coefficients dans $\mathbb{Z}$, on note que $\det(A)\det(A^{-1}) = 1$ alors que det(A) et $\det(A^{-1})$ sont des nombres entiers, encore par le corollaire 13.13. Ceci entraîne bien sûr que $\det(A) = \pm 1$. □

Autre application, nous pouvons maintenant répondre à la question posée dans l'exemple 4.29 :

Lemme 13.19 – *La fonction*

$$\begin{array}{ccc} GL_n(\mathbb{R}) & \longrightarrow & GL_n(\mathbb{R}) \\ A & \mapsto & A^{-1} \end{array}$$

est continue.

D'après l'expression pour A^{-1}, et le corollaire 13.13, c'est évident.

– Fin de la première lecture –

Nous allons nous tourner vers la démonstration du théorème 13.1. La partie « unicité » est en fait très simple, et en réalité nous l'avons presque déjà vue.

En effet, si $\phi\colon M_n(\mathbb{K}) \to \mathbb{K}$ vérifie les fameuses propriétés (1), (2), (3) et (4), alors pour calculer $\phi(A)$, on peut faire des opérations sur les lignes dont on sait précisément comment elles changent la quantité $\phi(A)$, et se ramener à une matrice échelonnée. Mais pour une matrice échelonnée, on constate que ϕ prend la valeur 0 s'il y a une ligne nulle, ou 1 sinon (voir la discussion dans l'exemple 13.4 et la démonstration de la proposition 13.5). Finalement, on n'a tout simplement pas le choix pour la valeur de $\phi(A)$. D'où l'unicité d'une fonction qui vérifierait les quatre propriétés.

Si ϕ ne vérifie pas la propriété (4), on pose $\alpha = \phi(\mathrm{Id})$. Si $\alpha \neq 0$, on travaille avec $\psi(A) = \phi(A)/\alpha$ qui vérifie les quatre propriétés, et donc $\psi(A) = \det(A)$ ce qui donne bien $\phi(A) = \alpha \det(A)$. Si par contre $\alpha = 0$ on voit facilement que $\phi(A) = 0$ pour toute matrice A en raisonnant comme ci-dessus ; là encore $\phi(A) = \alpha \det(A) = 0$.

Toute la difficulté réside dans l'existence de la fonction déterminant. Avec la formule pour le développement du déterminant, on pourrait donner une définition par récurrence, en partant des matrices 2×2. C'est possible, mais certaines choses vont être difficiles à montrer, comme par exemple le fait que $\det(A) = \det({}^tA)$. Nous allons utiliser une autre approche, qui est plus instructive.

Faisons une observation sur les matrices 3×3. En développant par les lignes ou les colonnes, on obtient

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -a_{13}a_{22}a_{31} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{11}a_{22}a_{33}\,.$$

On constate que le déterminant (s'il existe !) est une somme de termes de la forme $a_{1i_1}a_{2i_2}\cdots a_{ni_n}$ précédés de signes convenables. Même si ça paraît compliqué, nous allons partir de là pour donner une définition générale.

Définition 13.20 – Soit X un ensemble. Une *permutation* de X est une bijection $\sigma\colon X \to X$.

Nous allons particulièrement nous intéresser au cas $X = \{1, 2, 3, \dots, n\}$. On notera $\mathfrak{S}_n$ l'ensemble des permutations de ce X, et on appelle $\mathfrak{S}_n$ le n-ième *groupe symétrique*.

le symbole $\mathfrak{S}$ est un S majuscule en écriture gothique allemande.

On trouve bien d'autres notations pour le groupe symétrique, comme par exemple S_n ou Σ_n.

Exemple 13.21 – Nous allons avoir besoin d'une notation pour donner des exemples. Nous allons écrire

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix};$$

c'est-à-dire que, par exemple,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

est un raccourci pour désigner la fonction σ, définie de l'ensemble $\{1, 2, 3\}$ vers lui-même, telle que $\sigma(1) = 3$, $\sigma(2) = 2$ et $\sigma(3) = 1$.

Pour vérifier que σ est bien une bijection, et donc un élément de $\mathfrak{S}_3$, le plus simple est de constater que $\sigma(\sigma(i)) = i$ pour $i = 1, 2, 3$, donc $\sigma^{-1} = \sigma$.

Voyons un élément de $\mathfrak{S}_5$:

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix}.$$

Pour montrer que τ est une bijection, donnons directement sa réciproque :

$$\tau^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 2 & 5 & 4 \end{pmatrix}.$$

On vérifie que $\tau(\tau^{-1}(i)) = i$ et $\tau^{-1}(\tau(i)) = i$, ce qui justifie la notation τ^{-1}. Pour trouver ce τ^{-1}, la recette que nous venons d'appliquer est simple : pour chaque i on cherche le nombre j tel que $\tau(j) = i$, qui doit exister et être unique (sinon τ n'est pas une bijection) ; c'est ce nombre que l'on écrit en dessous de i, c'est-à-dire que $j = \tau^{-1}(i)$.

Voici une façon encore plus simple de le dire : échangeons les deux lignes de la matrice de τ, puis déplaçons les colonnes pour que sur la première ligne on ait $1, 2, 3, 4, 5$ dans cet ordre. On obtient la matrice de τ^{-1}.

Définition 13.22 – Soit $\sigma \in \mathfrak{S}_n$. Pour deux entiers distincts i et j entre 1 et n, on pose

$$q_{ij} = \frac{\sigma(j) - \sigma(i)}{j - i}.$$

On a $q_{ij} = q_{ji}$, de sorte que q_{ij} ne dépend que de la paire $\{i, j\}$. On pose ensuite

$$\varepsilon(\sigma) = \prod_{\{i,j\}} q_{ij}.$$

On appelle $\varepsilon(\sigma)$ la *signature* de σ.

Exemple 13.23 – Prenons

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \in \mathfrak{S}_3.$$

Les paires à considérer sont $\{1, 2\}, \{1, 3\}, \{2, 3\}$. On trouve

$$q_{12} = \frac{3-2}{2-1} = 1, \quad q_{13} = \frac{1-2}{3-1} = -\frac{1}{2}, \quad q_{23} = \frac{1-3}{3-2} = -2.$$

Finalement $\varepsilon(\sigma) = 1 \times -\frac{1}{2} \times (-2) = 1$.

Notons que la quantité q_{ij} est positive si $\sigma(i)$ et $\sigma(j)$ sont « dans le même ordre » que i et j, et négative sinon : on dit alors qu'il y a une *inversion* en i, j.

Lemme 13.24 – *Pour toute permutation σ, on a $\varepsilon(\sigma) = \pm 1$.*

Démonstration. Montrons que $|\varepsilon(\sigma)| = 1$. En posant $d_{ij} = |j - i|$, de sorte que $d_{ij} = d_{ji}$, on peut écrire $|\varepsilon(\sigma)| = N/D$ avec

$$N = \prod_{\{i,j\}} d_{\sigma(i)\sigma(j)} \quad \text{et} \quad D = \prod_{\{i,j\}} d_{ij}.$$

On doit donc montrer que $N = D$. Or c'est une évidence : dans les deux cas, il s'agit du produit de *tous* les nombres d_{ij} pour toutes les paires $\{i, j\}$. Voici une autre formulation : pour chaque terme d_{ij} du produit D, considérons $i' = \sigma^{-1}(i)$ et $j' = \sigma^{-1}(j)$. Alors N contient le terme $d_{\sigma(i')\sigma(j')} = d_{ij}$. Ainsi chaque terme du produit D correspond à un terme et un seul du produit N, et vice-versa, d'où $N = D$. □

Corollaire 13.25 – *Soit N le nombre d'inversions pour σ. Alors*

$$\varepsilon(\sigma) = (-1)^N.$$

Démonstration. Seul le signe de q_{ij} compte, dans le calcul de la signature $\varepsilon(\sigma)$, puisqu'en valeur absolue on sait déjà que $|\varepsilon(\sigma)| = 1$. Ce signe vaut (-1) si et seulement si on a une inversion en i, j, donc au total le signe est $(-1)^N$. □

Exemple 13.26 – Fixons i et j et considérons la *transposition* $\tau_{ij} \in \mathfrak{S}_n$, définie par $\tau_{ij}(i) = j$, $\tau_{ij}(j) = i$ et $\tau_{ij}(x) = x$ si $x \neq i, j$. En d'autres termes, τ_{ij} échange i et j et fixe les autres éléments. Par exemple dans $\mathfrak{S}_5$ on a

$$\tau_{24} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 2 & 5 \end{pmatrix}.$$

Calculons la signature de τ_{ij}, en supposant $i < j$. En comptant les inversions, on constate que l'on en obtient une pour la paire $\{i, j\}$, une pour chaque paire $\{i, x\}$ avec $i < x < j$, et une pour chaque paire $\{x, j\}$ avec $i < x < j$. On peut regrouper ces dernières inversions par deux en associant $\{i, x\}$ et $\{x, j\}$, ce qui rassemble un nombre pair de signes (-1) dans le calcul de $\varepsilon(\sigma)$. L'inversion pour $\{i, j\}$ reste seule, et au total

$$\varepsilon(\tau_{ij}) = -1.$$

Définition 13.27 – Soient σ et τ des éléments de $\mathfrak{S}_n$. La composition $\sigma \circ \tau$ définie par $\sigma \circ \tau(i) = \sigma(\tau(i))$ est encore un élément de $\mathfrak{S}_n$, que l'on va noter $\sigma\tau$ par simplicité. La composition est parfois appelée le *produit* de σ et τ.

Certains auteurs et certains logiciels préfèrent la notation $\tau\sigma$ au lieu de $\sigma\tau$. Attention !

Proposition 13.28 – *Pour $\sigma, \tau \in \mathfrak{S}_n$, on a*

$$\varepsilon(\sigma\tau) = \varepsilon(\sigma)\,\varepsilon(\tau).$$

Démonstration. Écrivons

$$\frac{\sigma\tau(j) - \sigma\tau(i)}{j - i} = \frac{\sigma\tau(j) - \sigma\tau(i)}{\tau(j) - \tau(i)} \cdot \frac{\tau(j) - \tau(i)}{j - i}.$$

Multiplions ces égalités pour toutes les paires $\{i, j\}$, on obtient

$$\varepsilon(\sigma\tau) = \left(\prod_{\{i,j\}} \frac{\sigma\tau(j) - \sigma\tau(i)}{\tau(j) - \tau(i)} \right) \varepsilon(\tau).$$

Il ne reste qu'à montrer que le produit entre parenthèses est effectivement $\varepsilon(\sigma)$, c'est-à-dire que

$$\prod_{\{i,j\}} \frac{\sigma\tau(j) - \sigma\tau(i)}{\tau(j) - \tau(i)} = \prod_{\{i,j\}} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Or chaque terme d'un de ces produits se retrouve une fois et une seule dans l'autre (comme dans la démonstration du lemme 13.24), donc ils sont bien égaux. □

Les groupes

C'est Évariste Galois qui le premier a utilisé la notion suivante : une partie $G \subset \mathfrak{S}_n$ est appelée un *groupe de permutations* lorsque (i) le produit $\sigma\tau$ de deux éléments $\sigma, \tau \in G$ vérifie encore $\sigma\tau \in G$, et (ii) si $\sigma \in G$ alors $\sigma^{-1} \in G$ (même si en fait on peut montrer que la condition (i) implique la (ii) automatiquement). Par exemple $G = \mathfrak{S}_n$ est un groupe de permutations, et l'ensemble A_n des éléments σ de $\mathfrak{S}_n$ vérifiant $\varepsilon(\sigma) = 1$ aussi (on l'appelle le « groupe alterné »).

Les travaux spectaculaires de Galois en théorie des corps ont montré la pertinence de ce concept.

Plus tard a emergé la définition de *groupe* (ou « groupe abstrait ») : il s'agit d'un ensemble G muni d'une multiplication associative, pour laquelle il existe un « élément neutre » e tel que $eg = ge = g$ pour chaque $g \in G$, et telle que chaque élément $g \in G$ possède un inverse $g^{-1} \in G$ qui satisfait $gg^{-1} = g^{-1}g = e$.

Nous avons vu de nombreux exemples. Le « groupe symétrique » $\mathfrak{S}_n$ est un groupe, la multiplication étant la composition des permutations, l'élément neutre étant l'identité (la permutation $x \mapsto x$). De même tout groupe de permutations est un groupe, comme on s'y attendait.

Également, l'ensemble $GL_n(\mathbb{K})$ des matrices inversibles à coefficients dans $\mathbb{K}$ est un groupe, la multiplication étant celle que nous connaissons sur les matrices, et l'élément neutre étant la matrice identité.

L'ensemble $\mathbb{Z}$ des entiers est un groupe lui aussi, mais attention : l'opération est l'addition, l'élément neutre est 0 ! Si on regarde la multiplication usuelle des entiers, avec 1 comme élément neutre, alors certains éléments de $\mathbb{Z}$ n'ont pas d'inverse dans $\mathbb{Z}$ (comme 2, son inverse $1/2$ n'est pas dans $\mathbb{Z}$). Par contre l'ensemble $\mathbb{Q}^\times$ des rationnels *non-nuls* est un groupe pour la multiplication usuelle, d'élément neutre 1. Et l'ensemble à deux éléments $\{\pm 1\}$ est aussi un groupe.

Un *homomorphisme de groupes* est une application $\phi\colon G \to H$, où G et H sont des groupes, qui vérifie $\phi(gh) = \phi(g)\phi(h)$. Ainsi nous venons de montrer dans ce chapitre que la signature est un homomorphisme $\varepsilon\colon \mathfrak{S}_n \to \{\pm 1\}$. Le fait que $\det\colon GL_n(\mathbb{K}) \to \mathbb{K}^\times$ est un homomorphisme doit également être frais dans votre mémoire.

La voici enfin :

Définition 13.29 – Soit $A = (a_{ij}) \in M_n(\mathbb{K})$. Son *déterminant* est

$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Nous allons montrer toutes les propriétés (attendues depuis l'énoncé du théorème 13.1) sous forme de lemmes.

Lemme 13.30 – *On a* $\det(A) = \det({}^tA)$.

Démonstration. On doit vérifier que

$$\sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}.$$

Fixons σ, et observons les nombres $a_{\sigma(i)i}$. En prenant $i = \sigma^{-1}(j)$, pour un entier j quelconque, on obtient $a_{\sigma(i)i} = a_{j\sigma^{-1}(j)}$; et en prenant le produit on a

$$a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} a_{2\sigma^{-1}(2)} \cdots a_{n\sigma^{-1}(n)}.$$

De plus, de la relation $\sigma\sigma^{-1} = \mathrm{Id}$ (la permutation identité), on déduit $\varepsilon(\sigma)\varepsilon(\sigma^{-1}) = 1$ et donc $\varepsilon(\sigma) = \varepsilon(\sigma^{-1})$.

Finalement, le terme correspondant à σ dans le membre de droite ci-dessus est précisément le terme correspondant à σ^{-1} dans le membre de gauche. Donc les sommes sont égales. □

Lemme 13.31 – *Si* A_1 *est obtenue à partir de* A *en multipliant une ligne par* λ*, alors* $\det(A_1) = \lambda \det(A)$.

Celui-ci est évident !

Lemme 13.32 – *Si* A_2 *est obtenue à partir de* A *en permutant deux lignes, alors* $\det(A_2) = -\det(A)$.

Démonstration. Soient k et ℓ les lignes qui sont permutées. Si $A = (a_{ij})$, alors $A_2 = (a_{\tau(i)j})$, où $\tau = \tau_{k\ell}$ est la transposition comme dans l'exemple 13.26. Le déterminant de cette matrice est

$$\det(A_2) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{\tau(1)\sigma(1)} a_{\tau(2)\sigma(2)} \cdots a_{\tau(n)\sigma(n)}.$$

Notons que $a_{\tau(i)\sigma(i)} = a_{j\sigma(\tau(j))}$ avec $j = \tau(i)$, ou ce qui revient au même $i = \tau(j)$. On en tire

$$a_{\tau(1)\sigma(1)} a_{\tau(2)\sigma(2)} \cdots a_{\tau(n)\sigma(n)} = a_{1\sigma(\tau(1))} a_{2\sigma(\tau(2))} \cdots a_{n\sigma(\tau(n))}.$$

Combinons ça avec $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau) = -\varepsilon(\sigma)$ encore d'après l'exemple 13.26. Finalement

$$\det(A_2) = -\sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma\tau)\, a_{1\sigma(\tau(1))} a_{2\sigma(\tau(2))} \cdots a_{n\sigma(\tau(n))}.$$

Il reste à observer que la somme ci-dessus est det(A) : en effet le terme correspondant à σ dans la définition de det(A) se retrouve dans cette somme correspondant à $\sigma\tau^{-1}$. □

Pour le reste des démonstrations, fixons une matrice A, et choisissons une ligne i. Pour $x_1, x_2, \ldots, x_n \in \mathbb{K}$, on va noter $f(x_1, \ldots, x_n)$ le déterminant de la matrice obtenue en remplaçant la ligne i de A par $(x_1, \ldots, x_n)$. Par exemple si $A = (a_{ij})$ est une matrice 3×3 et que l'on regarde la ligne 2, cela signifie que

$$f(x_1, x_2, x_3) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ x_1 & x_2 & x_3 \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

En particulier on a $\det(A) = f(a_{i1}, a_{i2}, \ldots, a_{in})$.

D'après la définition 13.29, il est bien clair qu'il existe des nombres que l'on va noter $\lambda_1, \lambda_2, \ldots, \lambda_n$ tels que

$$f(x_1, \ldots, x_n) = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n. \qquad (*)$$

En conséquence, on note que $f(x_1 + y_1, \ldots, x_n + y_n) = f(x_1, \ldots, x_n) + f(y_1, \ldots, y_n)$.

Lemme 13.33 – *Les formules de développement du déterminant par une ligne ou une colonne sont valides.*

Démonstration. Puisque $\det(A) = \det({}^tA)$, il suffit de montrer ceci pour les lignes. D'après la définition 13.10, nous devons donc montrer que $\lambda_j = (-1)^{i+j}\Delta_{ij}$, avec les notations ci-dessus.

Notons que $\lambda_j = f(0, \ldots, 0, 1, 0, \ldots, 0)$, avec le 1 en j-ième position, d'après (*). On peut donc voir λ_j comme le déterminant d'une certaine matrice ; en permutant i lignes et j colonnes, cette matrice devient

$$B = \left(\begin{array}{ccc|c} & M & & * \\ & & & \vdots \\ & & & * \\ \hline 0 & \cdots & 0 & 1 \end{array} \right)$$

où M est obtenue à partir de A en supprimant la ligne i et la colonne j. Les opérations sur les lignes et colonnes ont introduit le signe $(-1)^{i+j}$, c'est-à-dire $\lambda_j = (-1)^{i+j}\det(B)$. Nous allons montrer que $\det(B) = \det(M)$; par définition $\Delta_{ij} = \det(M)$, donc on aura bien établi que le déterminant λ_j est $(-1)^{i+j}\Delta_{ij}$.

Nous sommes donc ramenés à un cas très particulier de développement par une ligne, celui de montrer que $\det(B) = \det(M)$. Écrivons $B = (b_{k\ell})$, de sorte que

$$\det(B) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)}.$$

L'essentiel est que $b_{n,\ell} = 0$ pour $1 \le \ell < n$. Si σ est une permutation telle que $\sigma(n) \neq n$, on a un terme nul dans la somme ci-dessus ; autrement dit on peut se contenter de regarder les σ tels que $\sigma(n) = n$; et comme $b_{n,n} = 1$, il reste

$$\det(B) = \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(n) = n}} \varepsilon(\sigma) b_{1\sigma(1)} \cdots b_{n-1,\sigma(n-1)}.$$

Faisons l'observation suivante. Si $\sigma \in \mathfrak{S}_n$ vérifie $\sigma(n) = n$, alors on obtient une permutation de $\{1, 2, \ldots, n-1\}$ par restriction de σ. Si on note $\sigma' \in \mathfrak{S}_{n-1}$ la restriction de $\sigma \in \mathfrak{S}_n$, on obtient en fait une bijection entre $\mathfrak{S}_{n-1}$ et les permutations de $\mathfrak{S}_n$ qui « fixent » n. De plus, la signature de σ est égale à celle de σ′. Finalement

$$\det(B) = \sum_{\sigma' \in \mathfrak{S}_{n-1}} \varepsilon(\sigma') b_{1\sigma'(1)} \cdots b_{n-1,\sigma'(n-1)} = \det(M),$$

par définition. □

Lemme 13.34 – *Lorsque la matrice* A *possède deux lignes identiques, on a* $\det(A) = 0$.

Démonstration. En permutant ces deux lignes, on ne change pas A ; donc $\det(A) = -\det(A)$. On en déduit que $\det(A) = 0$.

Les plus observateurs auront noté que cet argument ne fonctionne que parce que $1 \neq -1$, ou encore $2 \neq 0$. Or il se peut très bien que $2 = 0$ si l'on travaille avec $\mathbb{K} = \mathbb{Z}/2\mathbb{Z}$, ce qui n'est pas exclu ! Pour couvrir ce cas, on peut faire une démonstration du lemme par récurrence, en partant des matrices 2×2 et en développant par une ligne à l'aide du lemme précédent. □

Voici finalement la dernière propriété :

Lemme 13.35 – *Si* A_3 *est obtenue à partir de* A *en ajoutant un multiple de la ligne* j *à la ligne* i*, alors* $\det(A_3) = \det(A)$.

Démonstration. On a

$$\begin{aligned} \det(A_3) &= f(a_{i1} + \lambda a_{j1}, a_{i2} + \lambda a_{j2}, \ldots, a_{in} + \lambda a_{jn}) \\ &= f(a_{i1}, a_{i2}, \ldots, a_{in}) + \lambda f(a_{j1}, a_{j2}, \ldots, a_{jn}) \\ &= \det(A) + 0. \end{aligned}$$

En effet $f(a_{j1}, a_{j2}, \ldots, a_{jn}) = 0$, puisque c'est le déterminant de la matrice obtenue en recopiant la ligne j dans la ligne i (et qui possède donc deux lignes identiques). □

Chapitre 14

Espaces vectoriels

On rappelle que la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$.

Au collège on vous a présenté les vecteurs, dans le cadre de la géométrie élémentaire dans le plan ou l'espace. Ces méthodes sont tellement efficaces que l'on souhaite les appliquer le plus largement possible, non seulement en « dimension » quelconque, mais également dans des cadres abstraits. Un « espace vectoriel » va être défini comme un ensemble sur lequel on peut faire ce type de géométrie.

Il se trouve que les calculs que nous allons être amenés à faire se ramènent presque tous à des opérations sur les matrices, que nous savons déjà exécuter. Ce chapitre présente une organisation abstraite de ces calculs, en quelque sorte. Au fur et à mesure de vos études en mathématiques, les espaces vectoriels vont prendre de plus en plus de place.

DÉFINITION 14.1 – Un *espace vectoriel* sur $\mathbb{K}$ est tout d'abord un ensemble E possédant un élément distingué noté 0_E ou simplement 0, que l'on appelle le *vecteur nul de* E ; on suppose ensuite que E est muni d'une opération d'addition

$$\begin{array}{ccc} E \times E & \longrightarrow & E \\ (u,v) & \mapsto & u+v \end{array}$$

et d'une opération

$$\begin{array}{ccc} \mathbb{K} \times E & \longrightarrow & E \\ (\lambda,u) & \mapsto & \lambda \cdot u \end{array}$$

satisfaisant les axiomes suivants :

(a) $u+v = v+u$,
(b) $0_E + u = u$,
(c) $(u+v)+w = u+(v+w)$,
(d) $\forall u\ \exists(-u)$ tel que $u+(-u) = 0_E$,
(e) $\lambda \cdot (u+v) = \lambda \cdot u + \lambda \cdot v$,
(f) $(\lambda+\mu)\cdot u = \lambda\cdot u + \mu \cdot u$,
(g) $1 \cdot u = u$,
(h) $(\lambda\mu)\cdot u = \lambda \cdot (\mu \cdot u)$,

pour $u,v,w \in E$ et $\lambda,\mu \in \mathbb{K}$.

EXEMPLE 14.2 – L'exemple le plus fondamental, sans conteste, est celui de $\mathbb{K}^n$. On identifie, comme d'habitude, les éléments de $\mathbb{K}^n$ avec les matrices-colonnes de $M_{n,1}(\mathbb{K})$, et les opérations sont celles que l'on connait bien :

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix},$$

et

$$\lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

Le vecteur nul est, bien sûr, celui dont toutes les coordonnées sont nulles. On vérifie sans problème que les axiomes (a-b-c-d-e-f-g-h) sont satisfaits, donc $\mathbb{K}^n$ est un espace vectoriel sur $\mathbb{K}$. En fait la proposition 12.7 nous indique que l'ensemble $M_{n,m}(\mathbb{K})$ des matrices est aussi un espace vectoriel.

EXEMPLE 14.3 – L'ensemble $\mathbb{K}[X]$ des polynômes à coefficients dans $\mathbb{K}$ est un espace vectoriel sur $\mathbb{K}$, la vérification des huit axiomes étant immédiate. Ici le « vecteur nul » est le polynôme nul, celui dont tous les coefficients sont nuls. C'est un premier pas vers l'abstraction : de par leurs bonnes propriétés, les polynômes peuvent être vus et manipulés comme des vecteurs !

EXEMPLE 14.4 – Soit A un ensemble quelconque. Notons $\mathcal{F}(A,\mathbb{K})$ l'ensemble des fonctions $A \to \mathbb{K}$. Si f et g sont de telles fonctions, on définit leur somme $f+g$ de la manière la plus simple, par la formule $(f+g)(x) = f(x)+g(x)$, pour $x \in A$. De même si $\lambda \in \mathbb{K}$, on définit $\lambda \cdot f$ par $(\lambda \cdot f)(x) = \lambda f(x)$.

On vérifie que $\mathcal{F}(A,\mathbb{K})$, avec ces opérations, est un espace vectoriel sur $\mathbb{K}$. Les fonctions sont donc aussi des vecteurs. Le « vecteur nul » dans cette situation est la fonction identiquement nulle, c'est-à-dire la fonction $f : A \to \mathbb{K}$ définie par $f(x) = 0$ pour tout $x \in A$.

La quasi-totalité des espaces vectoriels que nous allons rencontrer vont être bâtis à partir des trois exemples précédents, en ajoutant des conditions supplémentaires. Nous allons utiliser la notion suivante :

DÉFINITION 14.5 – Soient E un espace vectoriel et $F \subset E$ une partie non vide. On dit que F est un *sous-espace vectoriel* de E lorsque les deux conditions suivantes sont remplies : pour $u, v \in F$, on doit avoir $u + v \in F$, et pour $\lambda \in \mathbb{K}$ et $v \in F$, on doit avoir $\lambda \cdot v \in F$.

Il est clair qu'un sous-espace vectoriel est lui-même un espace vectoriel, et c'est notre principale source d'exemples. Pour certains étudiants (par exemple certains chimistes), la seule notion « au programme » est celle de sous-espace vectoriel de $\mathbb{R}^n$. Et pour tout le monde, c'est l'exemple à comprendre en premier.

EXEMPLE 14.6 – Voici un exemple générique de sous-espace de $\mathbb{K}^n$. Donnons-nous une matrice $A \in M_{m,n}(\mathbb{K})$ et considérons

$$E = \{v \in \mathbb{K}^n \mid Av = 0\}.$$

(Là encore on identifie les éléments de $\mathbb{K}^n$ avec des matrices-colonnes, donc le produit Av a un sens.) Alors E est un sous-espace de $\mathbb{K}^n$. En effet si u et v sont dans E, on a $Au = Av = 0$, donc $A(u + v) = Au + Av = 0$, donc $u + v \in E$. On vérifie de même que $A(\lambda v) = \lambda Av = 0$ si $Av = 0$, donc $\lambda v \in E$ si $v \in E$.

Par exemple, prenons

$$A = \begin{pmatrix} -3 & 1 & 2 \\ 7 & 0 & 8 \end{pmatrix} \in M_{2,3}(\mathbb{R}).$$

Définissons E comme ci-dessus, et prenons un élément

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

Alors $v \in E$ lorsque $Av = 0$, c'est-à-dire lorsque

$$\begin{cases} -3x + y + 2z = 0 \\ 7x + 8z = 0 \end{cases}.$$

On dit que ce sont les équations qui définissent E. On retiendra que les solutions d'un système linéaire, dont le « second membre » est 0, forment un espace vectoriel.

EXEMPLE 14.7 – On note $\mathbb{K}_n[X]$ l'ensemble des polynômes dont le degré est $\leq n$. Alors $\mathbb{K}_n[X]$ est un sous-espace vectoriel de l'espace $\mathbb{K}[X]$ (vérifiez-le).

Voyons un exemple plus abstrait.

EXEMPLE 14.8 – Soit $\mathcal{C}(I, \mathbb{R})$ l'ensemble des fonctions continues sur l'intervalle I. On a $\mathcal{C}(I, \mathbb{R}) \subset \mathcal{F}(I, \mathbb{R})$, et la proposition 4.11 nous dit que c'est un sous-espace vectoriel.

On peut remplacer « continue » par « dérivable », ou encore « paire », ou « impaire »... On peut même considérer

$$E = \{f \in \mathcal{F}(I, \mathbb{R}) \mid f \text{ est dérivable deux fois et } 3f'' - 5f' + f = 0\},$$

on vérifie que E est alors un sous-espace vectoriel de $\mathcal{F}(I, \mathbb{R})$.

voir les exercices 886, 888, 893, 5164, 5165

Pour décrire un sous-espace vectoriel, il est très commun de donner des équations comme dans l'exemple 14.6, mais il y a une autre méthode également utile.

DÉFINITION 14.9 – Soit E un espace vectoriel, et soient $e_1, e_2, \ldots, e_n$ des éléments de E. Une *combinaison linéaire* de ces éléments est une somme de la forme

$$\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n ,$$

avec $\lambda_i \in \mathbb{K}$.

L'ensemble des combinaisons linéaires de $e_1, e_2, \ldots, e_n$ est noté $\mathcal{V}ect(e_1, \ldots, e_n)$.

LEMME 14.10 – *L'ensemble $\mathcal{V}ect(e_1, \ldots, e_n)$ est un sous-espace vectoriel de* E.

Nous dirons de $\mathcal{V}ect(e_1, \ldots, e_n)$ que c'est l'espace *engendré* par les vecteurs $e_1, \ldots, e_n$.

Démonstration. On fait un calcul direct. Prenons

$$u = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n \in \mathcal{V}ect(e_1, \ldots, e_n)$$

et

$$v = \mu_1 e_1 + \mu_2 e_2 + \cdots + \mu_n e_n \in \mathcal{V}ect(e_1, \ldots, e_n),$$

on a alors

$$x + y = (\lambda_1 + \mu_1)e_1 + (\lambda_2 + \mu_2)e_2 + \cdots + (\lambda_n + \mu_n)e_n .$$

Ainsi $u + v \in \mathcal{V}ect(e_1, \ldots, e_n)$. De même on voit que λv appartient au vect si c'est le cas de v. □

EXEMPLE 14.11 – Prenons $E = \mathbb{R}^3$, puis

$$e_1 = \begin{pmatrix} 1 \\ 5 \\ -2 \end{pmatrix} \quad \text{et} \quad e_2 = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}.$$

Comment vérifier si un élément quelconque de $\mathbb{R}^3$, disons

$$v = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

appartient à $\mathcal{V}ect(e_1, e_2)$? Par définition $v \in \mathcal{V}ect(e_1, e_2)$ si et seulement s'il existe deux nombres λ_1 et λ_2 tels que $v = \lambda_1 e_1 + \lambda_2 e_2$. En écrivant ceci, on tombe sur le système suivant :

$$\begin{cases} \lambda_1 & + & 3\lambda_2 & = & x \\ 5\lambda_1 & & & = & y \\ -2\lambda_1 & - & 3\lambda_2 & = & z \end{cases}$$

Nous savons faire ; commençons par $L_2 \leftarrow L_2 - 5L_1$ et $L_3 \leftarrow L_3 + 2L_1$:

$$\begin{cases} \lambda_1 & + & 3\lambda_2 & = & x \\ & & -15\lambda_2 & = & y - 5x \\ & & 3\lambda_2 & = & z + 2x \end{cases}$$

Faisons $L_2 \leftarrow L_2 + 5L_3$, puis échangeons les deux dernières lignes :

$$\begin{cases} \lambda_1 & + & 3\lambda_2 & = & x \\ & & 3\lambda_2 & = & z + 2x \\ & & 0 & = & 5x + y + 5z \end{cases}$$

Pour que $v \in \mathcal{V}ect(e_1, e_2)$, il est donc nécessaire que $5x + y + 5z = 0$. Mais réciproquement, si $5x+y+5z = 0$, alors on peut résoudre le système (à savoir, $\lambda_2 = \frac{1}{3}(z+2x)$ et $\lambda_1 = x - 3\lambda_2 = -x - z$, mais peu importent ces valeurs). Donc finalement l'espace vectoriel $\mathcal{V}ect(e_1, e_2)$ est complètement décrit par l'équation $5x + y + 5z = 0$.

Il est important de savoir passer d'un espace « décrit comme un vect » à une description « par des équations » comme dans l'exemple 14.6. On peut toujours le faire sur le modèle du calcul ci-dessus. Plus loin nous verrons comment faire la transition inverse (vous pouvez déjà essayer d'imaginer la méthode).

Les descriptions par des équations permettent de vérifier rapidement si un élément donné appartient au sous-espace en question ; les descriptions par les vects permettent d'obtenir facilement des vecteurs appartenant au sous-espace.

Un espace vectoriel donné peut être décrit comme un vect de plusieurs façons, et nous allons nous attacher à trouver les meilleures familles de vecteurs, notamment celles contenant le plus petit nombre d'éléments. Commençons par donner une définition :

DÉFINITION 14.12 – Soient E un espace vectoriel et $e_1, e_2, \ldots, e_n$ une famille d'éléments de E. On dit que c'est une *famille génératrice de* E lorsque $E = \mathcal{V}ect(e_1, \ldots, e_n)$.

EXEMPLE 14.13 – Prenons $E = \mathbb{R}^2$ et

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Écrivons simplement

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = x e_1 + y e_2 .$$

On constate bien que tout vecteur de $\mathbb{R}^2$ peut s'écrire comme une combinaison linéaire de e_1 et e_2, donc e_1, e_2 est une famille génératrice de $\mathbb{R}^2$.

Il y en a d'autres, par exemple prenons

$$\varepsilon_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \varepsilon_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

La condition pour qu'un vecteur $\begin{pmatrix} x \\ y \end{pmatrix}$ appartienne à $\mathcal{V}ect(\varepsilon_1, \varepsilon_2)$ est l'existence de deux nombres λ_1 et λ_2 tels que

$$\begin{cases} 2\lambda_1 & + & \lambda_2 & = & x \\ 3\lambda_1 & + & \lambda_2 & = & y \end{cases}$$

Le déterminant du système est $2 \times 1 - 3 \times 1 = -1 \neq 0$, donc la matrice correspondante est inversible (ou si vous préférez, la matrice bien échelonnée correspondante est l'identité), donc le système a une solution unique. On peut si on le souhaite trouver les valeurs de λ_1 et λ_2, mais peu importe : de toute façon, nous savons que $\varepsilon_1, \varepsilon_2$ est une famille génératrice de $\mathbb{R}^2$.

Enfin, notons que la famille $e_1, e_2, \varepsilon_1, \varepsilon_2$, qui comporte 4 éléments, est également génératrice par définition. Et pour terminer, nous avons vu un exemple de famille qui n'est pas génératrice dans l'exemple 14.11, puisque le vect en question n'était pas $\mathbb{R}^3$ tout entier mais un sous-espace décrit par une certaine équation.

En fait dans le cas où $E = \mathbb{K}^n$, on peut se ramener à des calculs simples sur des matrices :

PROPOSITION 14.14 – *Soit $e_1, e_2, \ldots, e_m$ une famille de vecteurs de $\mathbb{K}^n$, et soit $A \in M_{n,m}(\mathbb{K})$ la matrice dont les colonnes sont les vecteurs e_i. Enfin soit E_A la matrice bien échelonnée associée.*

Alors $e_1, e_2, \ldots, e_m$ est une famille génératrice de $\mathbb{K}^n$ si et seulement si E_A comporte un pivot dans chaque ligne.

Remarquons que la condition revient à demander que E_A ne comporte pas de ligne nulle.

Démonstration. Par définition, la famille est génératrice si et seulement si pour tout $v \in \mathbb{K}^n$, il existe

$$\Lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$$

tel que $A\Lambda = v$.

Supposons que c'est le cas. Prenons une matrice inversible P telle que $PA = E_A$ (corollaire 12.22), et choisissons

$$v = P^{-1} u \quad \text{avec} \quad u = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Dans ce cas le système $A\Lambda = v$ équivaut, en multipliant par P, à $PA\Lambda = Pv$, soit $E_A \Lambda = u$. On constate que la dernière ligne de E_A ne peut pas être nulle, sinon $E_A\Lambda$ se terminerait aussi par une ligne nulle, et ce n'est pas le cas de u. Donc aucune ligne de E_A n'est nulle, étant donné qu'elle est bien échelonnée.

Montrons la réciproque, et supposons que E_A n'a pas de ligne nulle. Nous devons montrer que le système $A\Lambda = v$ a des solutions quel que soit v, ou ce qui revient au même en multipliant par P, que $E_A \Lambda = Pv$ possède toujours des solutions. Notons que dans ce cas, puisque E_A est bien échelonnée avec n lignes non-nulles, de taille $n \times m$, on peut l'obtenir à partir de la matrice identité de taille $n \times n$ en rajoutant des colonnes à droite. Mais alors si l'on prend

$$\Lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

on obtient

$$E_A \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \text{les } n \text{ premières lignes de } \Lambda ,$$

et le système $E_A \Lambda = Pv$ possède certainement des solutions, puisqu'il s'écrit en fait λ_i = le coefficient sur la ligne i de Pv, pour $1 \leq i \leq n$. □

La remarque suivante est très utile :

COROLLAIRE 14.15 – *Si $e_1, e_2, \ldots, e_m$ est une famille génératrice de $\mathbb{K}^n$, alors $m \geq n$.*

De plus, si $m = n$, alors la famille est génératrice si et seulement si la matrice A est inversible.

Démonstration. Si $m < n$ la matrice échelonnée E_A, ayant plus de lignes que de colonnes, est certaine d'avoir une ligne nulle, donc la famille ne pourrait pas être génératrice d'après la proposition. Donc $m \geq n$.

Si $m = n$, la matrice E_A est carrée ; elle ne possède pas de ligne nulle exactement lorsqu'elle vaut l'identité, puisqu'elle est bien échelonnée. D'après la proposition 12.19, ceci équivaut à l'inversibilité de A. □ *voir les exercices 900, 908, 914*

Définition 14.16 – Soient E un espace vectoriel et $e_1, e_2, \ldots, e_n$ une famille d'éléments de E. On dit que c'est une *famille libre* lorsque l'équation

$$\lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0,$$

avec $\lambda_i \in \mathbb{K}$, ne possède qu'une seule solution, à savoir $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.

Nous allons voir que le concept de famille libre est en un certain sens le « dual » du concept de famille génératrice. Rapidement, nous verrons que les familles qui sont à la fois libres et génératrices sont particulièrement intéressantes. Mais commençons par des exemples.

Exemple 14.17 – Prenons $E = \mathbb{R}^2$, puis

$$e_1 = \begin{pmatrix} -5 \\ 1 \end{pmatrix} \quad \text{et} \quad e_2 = \begin{pmatrix} 3 \\ 7 \end{pmatrix}.$$

Pour vérifier si la famille est libre, nous devons examiner l'équation $\lambda_1 e_1 + \lambda_2 e_2 = 0$, qui s'écrit comme le système

$$\begin{cases} -5\lambda_1 & + & 3\lambda_2 & = & 0 \\ \lambda_1 & + & 7\lambda_2 & = & 0 \end{cases}.$$

Le déterminant étant $-38 \neq 0$, le système a une solution unique, qui est bien sûr $\lambda_1 = \lambda_2 = 0$. Donc la famille est libre.

Si maintenant on pose $e_3 = \begin{pmatrix} -2 \\ 8 \end{pmatrix}$, la famille e_1, e_2, e_3 est-elle libre ? Le système devient

$$\begin{cases} -5\lambda_1 & + & 3\lambda_2 & - & 2\lambda_3 = & 0 \\ \lambda_1 & + & 7\lambda_2 & + & 8\lambda_3 = & 0 \end{cases}.$$

En faisant $L_1 \leftarrow L_1 + 5L_2$, puis en permutant les lignes, on obtient

$$\begin{cases} \lambda_1 & + & 7\lambda_2 & + & 8\lambda_3 = & 0 \\ & & 38\lambda_2 & + & 38\lambda_3 = & 0 \end{cases}.$$

Le système est échelonné, on prend λ_3 comme paramètre, et on tire $\lambda_1 = \lambda_2 = -\lambda_3$. En particulier, on a la solution $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = -1$, et d'ailleurs on vérifie effectivement que $e_1 + e_2 - e_3 = 0$. La famille n'est donc pas libre. (Ceux d'entre vous qui auraient repéré que $e_1 + e_2 - e_3 = 0$ peuvent simplement faire cette remarque, et il est alors établi que la famille n'est pas libre.)

Exemple 14.18 – Voyons un exemple plus abstrait. On prend $E = \mathcal{F}(\mathbb{R}, \mathbb{R})$, l'espace vectoriel de toutes les fonctions $\mathbb{R} \to \mathbb{R}$, et on essaie la famille consituée par $e_1 = \cos$ (la fonction cosinus) et $e_2 = \sin$. Cette famille est-elle libre ?

On doit étudier l'équation $\lambda_1 e_1 + \lambda_2 e_2 = 0$. C'est une égalité de fonctions, et en particulier le 0 désigne la fonction nulle ; c'est-à-dire que l'équation est vraiment

$$\forall x \in \mathbb{R}, \qquad \lambda_1 \cos(x) + \lambda_2 \sin(x) = 0,$$

les inconnues étant λ_1 et λ_2. Or pour $x = 0$ on trouve $\lambda_1 = 0$ et pour $x = \frac{\pi}{2}$ on trouve $\lambda_2 = 0$, donc la famille est bien libre.

Pour étudier les familles libres dans $\mathbb{K}^n$, on dispose du résultat suivant, qu'il est instructif de comparer à la proposition 14.14.

Proposition 14.19 – *Soit $e_1, e_2, \ldots, e_m$ une famille de vecteurs de $\mathbb{K}^n$, et soit $A \in M_{n,m}(\mathbb{K})$ la matrice dont les colonnes sont les vecteurs e_i. Enfin soit E_A la matrice bien échelonnée associée.*

Alors $e_1, e_2, \ldots, e_m$ est une famille libre de $\mathbb{K}^n$ si et seulement si E_A comporte un pivot dans chaque colonne.

Démonstration. Pour vérifier si la famille est libre, on doit étudier le système $A\Lambda = 0$, avec

$$\Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}.$$

Il possède les mêmes solutions que le système échelonné $E_A\Lambda = 0$. À ce stade on doit se rappeler que les inconnues qui vont servir de paramètres dans l'écriture des solutions sont celles qui correspondent aux colonnes dans lesquelles il n'y a pas de pivot (relire au besoin l'exemple 12.13).

Ainsi la famille est libre $\iff$ le système n'a qu'une solution $\iff$ il n'y a pas de paramètres $\iff$ il y a un pivot dans chaque colonne. □

Corollaire 14.20 – *Si $e_1, e_2, \ldots, e_m$ est une famille libre de $\mathbb{K}^n$, alors $m \leq n$.*

De plus, si $m = n$, alors la famille est libre si et seulement si la matrice A est inversible.

Démonstration. Si $m > n$ la matrice échelonnée E_A, ayant plus de colonnes que de lignes, est certaine d'avoir une colonne sans pivot, donc la famille ne pourrait pas être libre. Donc $m \leq n$.

Si $n = m$, la matrice E_A est carrée ; elle possède un pivot dans chaque colonne exactement lorsqu'elle vaut l'identité, puisqu'elle est bien échelonnée. D'après la proposition 12.19, ceci équivaut à l'inversibilité de A. □

En comparant ce dernier résultat avec le corollaire 14.15, on constate que

Corollaire 14.21 – *Considérons une famille comportant précisément n vecteurs dans $\mathbb{K}^n$. Alors elle est libre $\iff$ elle est génératrice $\iff$ la matrice A est inversible.*

Autre observation simple : une famille à la fois libre et génératrice de $\mathbb{K}^n$ doit comporter n vecteurs, ni plus ni moins. Ces phénomènes sont généraux dans les espaces vectoriels, et nous allons tout de suite le montrer. *voir les exercices 2781, 5567*

Définition 14.22 – Lorsqu'une famille est à la fois libre et génératrice, on dit que c'est une *base* de l'espace vectoriel considéré.

Exemple 14.23 – Dans $\mathbb{K}^n$, nous venons juste de voir qu'une famille $e_1, e_2, \ldots, e_m$ ne peut pas être une base si $m \neq n$; si $n = m$ la famille est une base exactement lorsque la matrice $n \times n$ dans laquelle on a rangé ces vecteurs en colonnes est inversible.

La *base canonique* de $\mathbb{K}^n$ est celle pour laquelle la matrice en question est l'identité. En clair

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Exemple 14.24 – Considérons $E = \mathbb{K}_n[X]$, l'espace vectoriel des polynômes de degré $\leq n$. Posons $e_i = X^i$, pour $0 \leq i \leq n$. Tout polynôme s'écrit comme combinaison linéaire des puissances de X, donc la famille est génératrice ; de plus si

$$\lambda_0 e_0 + \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n = 0 = \lambda_0 + \lambda_1 X + \lambda_2 X^2 + \cdots + \lambda_n X^n,$$

alors on a $\lambda_0 = \lambda_1 = \cdots = \lambda_n = 0$ (par définition même de ce qu'est le polynôme nul). Donc la famille est libre.

Finalement la famille $1, X, X^2, \ldots, X^n$ est une base, qu'on appelle encore la *base canonique* de $\mathbb{K}_n[X]$. Noter qu'elle comprend $n+1$ éléments.

Exemple 14.25 – Lorsqu'un sous-espace de $\mathbb{K}^n$ est donné par des équations (sur le modèle de l'exemple 14.6), on peut facilement en trouver une base. Considérons donc

$$E = \{v \in \mathbb{K}^n \mid Av = 0\},$$

où A est une matrice. On peut échelonner la matrice sans changer les solutions de $Av = 0$, donc nous allons supposer que A est échelonnée et donner la méthode sur un exemple. Prenons

$$A = \begin{pmatrix} 1 & 4 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad v = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix},$$

alors en étudiant $Av = 0$ on constate que l'ensemble des solutions est

$$E = \left\{ y \begin{pmatrix} -4 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 3 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ avec } y, t \in \mathbb{K} \right\}.$$

(Pour les calculs intermédiaires, reprendre l'exemple 12.13.) Prenons les notations

$$e_1 = \begin{pmatrix} -4 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 3 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

alors nous venons d'écrire que $E = \mathcal{V}ect(e_1, e_2)$, donc e_1, e_2 est une famille génératrice de E. On peut vérifier directement que la famille est libre, puisque l'équation $\lambda_1 e_1 + \lambda_2 e_2 = 0$ donne

$$\begin{pmatrix} -4\lambda_1 + 3\lambda_2 \\ \lambda_1 \\ 0 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

d'où $\lambda_1 = \lambda_2 = 0$. La famille est bien libre et c'est donc une base de E.

Ce n'est pas un hasard : lorsqu'on écrit les solutions de la manière décrite dans l'exemple 12.13, les vecteurs que l'on obtient forment toujours une base. On peut le vérifier rapidement dans chaque cas.

À la fin du chapitre nous verrons comment trouver une base d'un sous-espace présenté comme un vect. En attendant, vous pourriez écrire des équations pour l'espace et procéder comme ci-dessus (mais la méthode que nous verrons est plus efficace).

voir les exercices 984, 997, 1013

L'intérêt des bases est de permettre l'utilisation de *coordonnées*, de la façon suivante. Soit $e_1, e_2, \dots, e_n$ une base de l'espace vectoriel E, et soit $v \in E$ un vecteur quelconque. La famille étant génératrice, on peut trouver des nombres λ_i tels que

$$v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n .$$

La famille étant libre, on peut voir que cette écriture est en fait *unique* : en effet, si on a également

$$v = \mu_1 e_1 + \mu_2 e_2 + \cdots + \mu_n e_n ,$$

alors en faisant la différence on obtient

$$v - v = 0 = (\lambda_1 - \mu_1)e_1 + (\lambda_2 - \mu_2)e_2 + \cdots + (\lambda_n - \mu_n)e_n .$$

Puisque la famille est libre, on doit avoir $\lambda_i - \mu_i = 0$ et donc $\lambda_i = \mu_i$.

Définition 14.26 – Soit $\mathscr{B} = e_1, e_2, \dots, e_n$ une base de l'espace vectoriel E. Si $v \in E$, les nombres $\lambda_1, \lambda_2, \dots, \lambda_n$ tels que

$$v = \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_n e_n$$

sont appelés les *coordonnées de v dans la base $\mathscr{B}$*. On notera

$$_{\mathscr{B}}[v] = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathbb{K}^n .$$

Lorsque la base $\mathscr{B}$ sera évidente d'après le contexte on écrira tout simplement $[v]$.

Exemple 14.27 – Soit $\mathscr{B} = e_1, e_2$ la base de $\mathbb{R}^2$ donnée par

$$e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{et} \quad e_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} .$$

C'est bien une base, puisque si nous mettons ces vecteurs en colonnes dans

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

alors $\det(A) = -2 \neq 0$. Prenons maintenant un vecteur quelconque de $\mathbb{R}^2$, disons

$$v = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

Pour trouver ses coordonnées λ_1, λ_2 dans la base $\mathscr{B}$ nous devons résoudre $\lambda_1 e_1 + \lambda_2 e_2 = v$, ce qui revient à

$$A\Lambda = v \quad \text{avec} \quad \Lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} .$$

Puisque A est inversible on a

$$\Lambda = A^{-1} v = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}x_1 + \frac{1}{2}x_2 \\ \frac{1}{2}x_1 - \frac{1}{2}x_2 \end{pmatrix} .$$

Finalement

$$_{\mathscr{B}}[v] = \begin{pmatrix} \frac{1}{2}x_1 + \frac{1}{2}x_2 \\ \frac{1}{2}x_1 - \frac{1}{2}x_2 \end{pmatrix} .$$

Par contre si nous appelons $\mathscr{C}$ la base canonique (comme dans l'exemple 14.23), alors on a tout simplement

$$_{\mathscr{C}}[v] = v .$$

En effet le calcul est le même que ci-dessus, mais cette fois A = Id ; ou encore, cela découle du calcul suivant :

$$v = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

Exemple 14.28 – Considérons maintenant $E = \mathbb{K}_n[X]$ et sa base canonique, c'est-à-dire $\mathscr{B} = 1, X, X^2, \dots, X^n$. Si on prend un polynôme $P \in E$ et qu'on l'écrit

$$P = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n ,$$

voir les exercices 979, 981, 3317

alors par définition

$$_{\mathscr{B}}[P] = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{K}^{n+1} .$$

Les coordonnées vont nous permettre de ramener de nombreuses questions abstraites sur un espace vectoriel E à des questions sur $\mathbb{K}^n$, que l'on sait traiter. Considérons par exemple :

Proposition 14.29 – *Soit E un espace vectoriel, et soit $\mathscr{B} = e_1, e_2, \dots, e_n$ une base. Écrivons $[v]$ pour $_{\mathscr{B}}[v]$. Alors*

1. $[u + v] = [u] + [v]$,
2. $[\lambda v] = \lambda [v]$,
3. *si $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ est une famille de vecteurs de E, alors elle est libre si et seulement si la famille $[\varepsilon_1], [\varepsilon_2], \dots, [\varepsilon_m]$ de vecteurs de $\mathbb{K}^n$ est libre.*
4. *idem avec « génératrice » au lieu de « libre ».*

La démonstration est extrêmement facile ; elle vous est laissée à titre d'exercice important.

Voici une première application. Le théorème suivant est le plus important du chapitre pour l'instant.

Théorème 14.30 – *Soit E un espace vectoriel, muni d'une base $\mathscr{B} = e_1, e_2, \dots, e_n$, et soit $\mathscr{F} = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ une famille de vecteurs de E.*

1. *Si $\mathscr{F}$ est génératrice, alors $m \geq n$.*
2. *Si $\mathscr{F}$ est libre, alors $m \leq n$.*
3. *Si $\mathscr{F}$ est une base, alors $m = n$.*

Réciproquement, si $m = n$, alors $\mathscr{F}$ est génératrice $\Longleftrightarrow$ $\mathscr{F}$ est libre $\Longleftrightarrow$ $\mathscr{F}$ est une base.

Démonstration. Elle est très simple. En effet nous avons démontré tout ceci dans le cas où $E = \mathbb{K}^n$: voir le corollaire 14.15 pour le (1), le corollaire 14.20 pour le (2) ; le (3) est alors évident, et la réciproque est également indiquée dans ces corollaires.

Pour le cas général, on utilise tout simplement la proposition précédente, qui nous ramène à $\mathbb{K}^n$. □

Nous constatons que *toutes les bases d'un espace vectoriel ont le même nombre d'éléments*. Ce nombre porte un nom :

Définition 14.31 – La *dimension* d'un espace vectoriel est le nombre de vecteurs dans une base quelconque. La dimension de E est notée $\dim E$.

Exemple 14.32 – La dimension de $\mathbb{K}^n$ est n : prendre la base canonique.

Exemple 14.33 – La dimension de $\mathbb{K}_n[X]$ est $n + 1$ (attention !), là encore voir la base canonique $1, X, X^2, \dots, X^n$.

Exemple 14.34 – L'espace vectoriel $\mathbb{K}[X]$ ne possède pas de base finie : en effet si $P_1, P_2, \dots, P_n$ est une famille finie de polynômes, alors en prenant $N = \sup\{\deg(P_i)\}$, on voit facilement que tout polynôme dans $\mathcal{Vect}(P_1, P_2, \dots, P_n)$ est de degré $\leq N$. En particulier $\mathcal{Vect}(P_1, \dots, P_n)$ n'est pas $\mathbb{K}[X]$ tout entier.

Dans ce livre nous ne parlerons pas de familles infinies de vecteurs. Ceci dit, il existe des définitions que vous pouvez imaginer de « famille libre », « famille génératrice » et « base » ayant éventuellement un nombre infini de vecteurs. Avec ces définitions, on montre que la famille infinie $1, X, X^2, \dots, X^k, \dots$ est une base de $\mathbb{K}[X]$.

Quoi qu'il en soit, nous dirons qu'un espace vectoriel est de *dimension finie* lorsqu'il possède une base finie. Ce n'est pas le cas de $\mathbb{K}[X]$, qui est de dimension infinie.

Exemple 14.35 – Voici une application célèbre. Prenons $E = \mathbb{R}_n[X]$, qui est de dimension $n + 1$. Choisissons un nombre $x_0 \in \mathbb{R}$, et considérons

$$e_i = (X - x_0)^i \quad \text{pour} \quad 0 \leq i \leq n .$$

Montrons que cette famille est libre : l'équation $\lambda_0 e_0 + \lambda_1 e_1 + \cdots + \lambda_n e_n = 0$ s'écrit

$$\lambda_0 + \lambda_1 (X - x_0) + \lambda_2 (X - x_0)^2 + \cdots + \lambda_n (X - x_0)^n = 0 .$$

Le terme en X^n dans le membre de gauche est $\lambda_n X^n$, donc $\lambda_n = 0$. Mais alors le terme en X^{n-1} dans le membre de gauche est $\lambda_{n-1} X^{n-1}$, donc $\lambda_{n-1} = 0$. De proche en proche, on en déduit que $\lambda_n = \lambda_{n-1} = \cdots = \lambda_1 = \lambda_0 = 0$. Donc la famille est libre.

Cette famille ayant $n + 1$ éléments, c'est en fait une base d'après le théorème. Autre argument possible pour montrer la même chose : écrivons $[e_i]$ pour les coordonnées de e_i dans la base canonique, alors il suffit de montrer que la famille $[e_i]$ est une base de $\mathbb{K}^{n+1}$ (cf proposition 14.29) ; or la matrice obtenue en mettant ces vecteurs en colonnes est triangulaire supérieure avec des 1 sur la diagonale, donc son déterminant est 1 et donc elle est inversible, ce qui établit que les vecteurs forment une base.

Que l'on prenne un argument ou l'autre, essayez d'apprécier les efforts qui nous sont économisés : montrer directement que la famille est génératrice par un calcul naïf serait bien pénible.

On en déduit que tout polynôme $P \in \mathbb{R}_n[X]$ peut s'écrire de manière unique sous la forme

$$P = \lambda_0 + \lambda_1 (X - x_0) + \lambda_2 (X - x_0)^2 + \cdots + \lambda_n (X - x_0)^n . \qquad (*)$$

Sachant que cette écriture existe, il est maintenant facile de calculer les nombres λ_i. En évaluant en $X = x_0$, on trouve déjà $\lambda_0 = P(x_0)$. Prenons maintenant la dérivée :

$$P' = \lambda_1 + 2\lambda_2 (X - x_0) + \cdots + n\lambda_n (X - x_0)^{n-1} .$$

On en tire $\lambda_1 = P'(x_0)$. En dérivant une deuxième fois on voit $2\lambda_2 = P''(x_0)$, puis $6\lambda_3 = P^{(3)}(x_0)$, et par récurrence on montre (faites-le) que $k!\,\lambda_k = P^{(k)}(x_0)$.

L'égalité $(*)$ s'appelle la *formule de Taylor*, qu'on écrit donc

$$P(X) = \sum_{k=0}^{\deg(P)} \frac{P^{(k)}(x_0)}{k!} (X - x_0)^k .$$

Dans le chapitre 7, nous montrons que cette formule (exacte !) peut se généraliser, de manière approchée, pour toute fonction suffisamment dérivable plutôt qu'un polynôme. Mais l'argument ci-dessus est purement algébrique, et s'appliquerait de la même manière aux polynômes dans $\mathbb{K}[X]$ pour toutes sortes de choix de $\mathbb{K}$. Tout juste faut-il donner une définition, algébrique elle aussi, de la dérivée d'un polynôme de $\mathbb{K}[X]$, et vérifier que les propriétés attendues sont bien satisfaites.

– Fin de la première lecture –

Nous avons vu que les bases sont très utiles pour étudier les espaces vectoriels, mais qu'il n'existe pas toujours de base finie (cf exemple 14.34). Nous aurions bien besoin de critères faciles pour garantir l'existence de bases, et c'est le théorème suivant qui va en donner. Commençons par un lemme très simple.

LEMME 14.36 – *Soit* E *un espace vectoriel, soit* $e_1, \ldots, e_n$ *une famille libre de* E*, et soit* $v \in \mathrm{E}$ *tel que la famille* $e_1, e_2, \ldots, e_n, v$ *n'est pas libre.*

Alors $v \in \mathcal{V}ect(e_1, \ldots, e_n)$.

Démonstration. Il existe une combinaison linéaire nulle

$$\lambda_1 e_1 + \cdots + \lambda_n e_n + \alpha v = 0\,,$$

dont les coefficients ne sont pas tous nuls. On doit donc avoir $\alpha \neq 0$: en effet dans le cas où $\alpha = 0$ on aurait

$$\lambda_1 e_1 + \cdots + \lambda_n e_n = 0$$

donc $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$ puisque la famille est libre, ce qui est une contradiction.

On peut donc écrire

$$v = -\frac{1}{\alpha}(\lambda_1 e_1 + \cdots + \lambda_n e_n)\,,$$

ce qui montre que $v \in \mathcal{V}ect(e_1, \ldots, e_n)$. □

THÉORÈME 14.37 (DE LA BASE INCOMPLÈTE) – *Soit* E *un espace vectoriel, soit* $\mathscr{L} = \ell_1, \ell_2, \ldots, \ell_n$ *une famille libre de vecteurs de* E*, et soit* $\mathscr{G} = g_1, g_2, \ldots, g_m$ *une famille génératrice de* E*. Alors on peut compléter* $\mathscr{L}$ *par des vecteurs de* $\mathscr{G}$ *pour en faire une base.*

Plus précisément, il existe des indices $i_1, i_2, \ldots, i_k$ *tels que*

$$\ell_1, \ell_2, \ldots, \ell_n, g_{i_1}, g_{i_2}, \ldots, g_{i_k}$$

est une base de E.

Démonstration. Considérons l'ensemble des entiers s tels qu'il existe des indices $i_1, i_2, \ldots, i_s$ pour lesquels

$$\ell_1, \ell_2, \ldots, \ell_n, g_{i_1}, g_{i_2}, \ldots, g_{i_s}$$

est libre ; notons S cet ensemble. Il est clair que de tels indices doivent être distincts (pourquoi ?), donc l'ensemble S est fini puisque tous ses éléments sont $\leq m$. Prenons $k = \sup \mathrm{S}$ (si $\mathrm{S} = \emptyset$ par contre, on prend $k = 0$). Par définition on a une famille libre

$$\mathscr{F} = \ell_1, \ell_2, \ldots, \ell_n, g_{i_1}, g_{i_2}, \ldots, g_{i_k}\,,$$

et nous allons montrer que c'est une base.

Prenons en effet un vecteur g_i de la famille $\mathscr{G}$. Si on l'ajoute à la famille $\mathscr{F}$, on obtient une famille de $n+k+1$ vecteurs ; par maximalité de k, cette famille ne peut pas être libre. D'après le lemme 14.36, on a donc $g_i \in \mathcal{V}ect(\mathscr{F})$.

C'est vrai pour tous les vecteurs de $\mathscr{G}$, donc $\mathcal{V}ect(\mathscr{G}) \subset \mathcal{V}ect(\mathscr{F})$. Mais $\mathcal{V}ect(\mathscr{G}) = \mathrm{E}$ puisque $\mathscr{G}$ est génératrice par définition, et donc $\mathcal{V}ect(\mathscr{F}) = \mathrm{E}$ également, c'est-à-dire que $\mathscr{F}$ est bien génératrice. □

COROLLAIRE 14.38 – *Si* E *possède une famille génératrice (finie), alors* E *possède une base (finie).*

Démonstration. C'est ce que dit le théorème dans le cas où $\mathscr{L}$ est la « famille vide » (vous pouvez vérifier que la démonstration du théorème est parfaitement adaptée au cas où il n'y a aucun vecteur dans $\mathscr{L}$). □

Voici une autre conséquence (en toute rigueur c'est surtout une conséquence du lemme 14.36).

COROLLAIRE 14.39 – *Soit* E *un espace vectoriel de dimension finie, et soit* $\mathrm{F} \subset \mathrm{E}$ *un sous-espace. Alors* F *est de dimension finie et* $\dim \mathrm{F} \leq \dim \mathrm{E}$.

De plus, on a une équivalence $\mathrm{F} = \mathrm{E} \Longleftrightarrow \dim \mathrm{F} = \dim \mathrm{E}$.

Démonstration. Si $\mathscr{L}$ est une famille libre de F, alors c'est aussi une famille libre de E. Donc $\mathscr{L}$ comprend moins de n éléments, où $n = \dim \mathrm{E}$ (théorème 14.30).

Prenons maintenant une famille libre $\mathscr{L}$ de F ayant le plus grand nombre d'éléments, disons $\mathscr{L} = \ell_1, \ldots, \ell_m$, avec donc $m \leq n$. Alors par le lemme 14.36, on voit que tout $f \in \mathrm{F}$ appartient à $\mathcal{V}ect(\mathscr{L})$ (puisque la famille $\ell_1, \ldots, \ell_m, f$ ne peut pas être libre, par maximalité de m). Cette famille est donc une base, et $\dim \mathrm{F} = m \leq \dim \mathrm{E}$.

Si on a en fait $\dim \mathrm{F} = \dim \mathrm{E}$, alors prenons une base $\mathscr{L} = \ell_1, \ell_2, \ldots, \ell_n$ de F ; c'est une famille libre d'éléments de E, comprenant $n = \dim \mathrm{E}$ vecteurs, donc c'est une base de E d'après le théorème 14.30. Ainsi $\mathrm{F} = \mathcal{V}ect(\mathscr{L}) = \mathrm{E}$. □

DÉFINITION 14.40 – Soit $A \in M_{n,m}(\mathbb{K})$. On note $Vect(A)$ le sous-espace de $\mathbb{K}^n$ engendré par les colonnes de A. La dimension de $Vect(A)$ est appelée le *rang* de la matrice A.

Nous allons voir comment calculer le rang. Du même coup, nous verrons comment trouver une base d'un espace vectoriel donné comme un vect dans $\mathbb{K}^n$. En fait, à l'aide de la théorie développée dans ce chapitre, nous allons donner *deux* méthodes profondément différentes ; le fait qu'elles donnent le même résultat est un théorème célèbre.

PROPOSITION 14.41 – *Le rang de A est le nombre de lignes non-nulles dans la matrice bien échelonnée* E_A.

Pour trouver une base de $Vect(A)$, *il suffit de prendre les colonnes de A qui correspondent aux pivots.*

Attention à la dernière phrase. Elle signifie que si les pivots de E_A sont dans les colonnes $i_1, \ldots, i_r$, alors les colonnes de la matrice de départ A numérotées $i_1, \ldots, i_r$ forment une base de $Vect(A)$ (notez que le nombre de pivots est égal au nombre de lignes non-nulles, bien sûr). De nombreux étudiants font l'erreur de proposer les colonnes de E_A comme base.

Démonstration. On note $g_1, g_2, \ldots, g_m$ les colonnes de A, qui forment une famille $\mathscr{G}$ génératrice de $Vect(A)$. Notons $i_1, \ldots, i_r$ les numéros des colonnes de E_A qui contiennent les pivots, et soit $\mathscr{B} = g_{i_1}, \ldots, g_{i_r}$. Nous allons montrer que $\mathscr{B}$ est une base de $Vect(A)$, ce qui prouve les deux assertions du même coup.

Soit B la matrice dont les colonnes sont les vecteurs de $\mathscr{B}$; en d'autres termes B est formée des colonnes de A numérotées $i_1, \ldots, i_r$. Alors la matrice bien échelonnée E_B est elle aussi extraite de E_A en gardant les colonnes correspondantes. En particulier, E_B possède un pivot dans chaque colonne, par construction. D'après la proposition 14.19, la famille $\mathscr{B}$ est libre.

Par contre, prenons un indice i qui n'est pas dans la liste $i_1, \ldots, i_r$, et rajoutons le vecteur g_i à $\mathscr{B}$ (à sa place dans l'ordre). La matrice de cette nouvelle famille, disons C, est obtenue à partir de B en réinsérant la colonne de A correspondante. La même chose peut être dite de E_C, obtenue à partir de E_B en rajoutant une colonne de E_A. Cette colonne ne contient pas de pivot par construction, et c'est encore la proposition 14.19 qui nous permet de conclure que la famille obtenue en rajoutant g_i à $\mathscr{B}$ n'est pas libre. D'après le lemme 14.36, on a $g_i \in Vect(\mathscr{B})$.

Finalement, on a $Vect(A) = Vect(\mathscr{G}) \subset Vect(\mathscr{B}) \subset Vect(A)$, donc $Vect(\mathscr{B}) = Vect(A)$, et $\mathscr{B}$ est une base. □

On en déduit la chose suivante :

COROLLAIRE 14.42 – *On ne change pas le rang d'une matrice en faisant des opérations sur ses lignes.*

Démonstration. Si A′ est obtenue à partir de A par de telles opérations, on a $E_{A'} = E_A$ clairement. □

EXEMPLE 14.43 – Prenons

$$A = \begin{pmatrix} 7 & -1 & 4 & 0 \\ 1 & 2 & 8 & 51 \\ 5 & -5 & -12 & -102 \end{pmatrix}.$$

Après quelques opérations sur les lignes, nous obtenons la forme bien échelonnée :

$$E_A = \begin{pmatrix} 1 & 0 & 16/15 & 17/5 \\ 0 & 1 & 52/15 & 119/5 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Le rang de A est donc 2. Les pivots sont dans les colonnes 1 et 2 de E_A, donc on va prendre les colonnes 1 et 2 de A :

$$e_1 = \begin{pmatrix} 7 \\ 1 \\ 5 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -1 \\ 2 \\ -5 \end{pmatrix}.$$

La proposition nous dit que e_1, e_2 est une base de $Vect(A)$.

voir les exercices 3318, 980 (ignorer les questions sur $F \cap G$ et $F + G$)

Passons à la deuxième méthode : les colonnes vont remplacer les lignes. Les choses se passent maintenant dans l'ordre inverse, car le résultat suivant est assez évident.

LEMME 14.44 – *On ne change pas le rang d'une matrice A en faisant des opérations sur les colonnes. En fait on ne change même pas* $Vect(A)$.

Démonstration. Si A′ est obtenue à partir de A par de telles opérations, chaque colonne de A′ est visiblement dans $Vect(A)$. Ainsi $Vect(A') \subset Vect(A)$. Mais bien sûr on peut retrouver A en faisant des opérations sur les colonnes de A′, donc de la même manière on a $Vect(A) \subset Vect(A')$. □

Dans l'énoncé suivant, on va dire qu'une matrice est *échelonnée en colonnes* lorsque c'est la transposée d'une matrice échelonnée. En d'autres termes, reprenez la définition de matrice échelonnée et remplacez « ligne » par « colonne ». En faisant des opérations sur les colonnes d'une matrice A, on peut la mettre sous une forme unique bien échelonnée en colonnes : pour s'en assurer, il suffit d'observer que cela revient à mettre la transposée ${}^tA$ sous forme bien échelonnée en faisant des opérations sur les lignes.

PROPOSITION 14.45 – *Le rang de A est le nombre de colonnes non-nulles dans la matrice bien échelonnée en colonnes associée à A.*

Pour trouver une base de $Vect(A)$, *il suffit de prendre les colonnes non-nulles de cette matrice échelonnée en colonnes.*

Démonstration. Soit B la matrice bien échelonnée en colonnes obtenue à partir de A. D'après le lemme $Vect(B) = Vect(A)$. Si $g_1, g_2, \ldots, g_r$ sont les colonnes non-nulles de B, il est clair que $Vect(g_1, \ldots, g_r) = Vect(B)$, donc il suffit de s'assurer que c'est une famille libre. Or les pivots étant seuls dans leurs lignes, c'est clair (voir l'exemple). □

EXEMPLE 14.46 – Reprenons le même exemple, c'est-à-dire :

$$A = \begin{pmatrix} 7 & -1 & 4 & 0 \\ 1 & 2 & 8 & 51 \\ 5 & -5 & -12 & -102 \end{pmatrix}.$$

En faisant des opérations sur les colonnes, on peut mettre A sous la forme

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & -2 & 0 & 0 \end{pmatrix},$$

et cette matrice est « bien échelonnée en colonnes ». On voit de nouveau que le rang est 2. Cette fois-ci, la proposition nous dit de prendre

$$\varepsilon_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \varepsilon_2 = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}.$$

La famille $\varepsilon_1, \varepsilon_2$ est une base de $Vect(A)$.

En comparant les deux méthodes, il vient le résultat suivant, qui est loin d'être évident si on part de la définition :

THÉORÈME 14.47 – *Le rang d'une matrice est égal au rang de sa transposée.*

Démonstration. Le rang de A est le nombre de lignes non-nulles dans E_A, qui est égal au nombre de colonnes non-nulles dans ${}^tE_A$. Or ${}^tE_A$ est bien échelonnée en colonnes, et obtenue à partir de ${}^tA$ en faisant des opérations sur les colonnes, donc le nombre de ses colonnes non-nulles est bien le rang de ${}^tA$. □

reprendre les exercices précédents avec le deuxième méthode

Chapitre 15

Applications linéaires

On rappelle que la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$.

Définition 15.1 – Soient E et F des espaces vectoriels sur $\mathbb{K}$. Une *application linéaire* est une fonction $f\colon E \to F$ telle que $f(u+v) = f(u) + f(v)$ pour $u, v \in E$, et telle que $f(\lambda v) = \lambda f(v)$ pour tout $\lambda \in \mathbb{K}$ et tout $v \in E$.

Notons que dans le cas où E = F, on dit parfois que f est un *endomorphisme de* E (pour dire que c'est une application linéaire E → E).

Exemple 15.2 – L'exemple le plus simple s'obtient en choisissant une matrice $A \in M_{m,n}(\mathbb{K})$. On définit alors une application

$$f\colon \mathbb{K}^n \longrightarrow \mathbb{K}^m, \quad v \mapsto f(v) = Av.$$

(Comme d'habitude, les vecteurs sont vus comme des matrices-colonnes.) On vérifie très simplement que $f(u+v) = A\cdot(u+v) = Au + Av = f(u) + f(v)$, et $f(\lambda v) = A\cdot(\lambda v) = \lambda Av = \lambda f(v)$. C'est donc bien une application linéaire.

Ainsi l'application $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ définie par

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -5x+7y-z \\ x-y+19z \end{pmatrix} = \begin{pmatrix} -5 & 7 & -1 \\ 1 & -1 & 19 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

est linéaire.

Exemple 15.3 – Prenons E = F = $\mathbb{C}$, qui est un espace vectoriel de dimension 1 sur $\mathbb{K} = \mathbb{C}$. Fixons un nombre réel θ. L'application $f(z) = e^{i\theta}z$ est alors linéaire, exactement comme dans l'exemple précédent.

On peut aussi identifier $\mathbb{C}$ avec $\mathbb{R}^2$ de la manière habituelle ; on le voit alors comme un espace vectoriel de dimension 2 sur $\mathbb{K} = \mathbb{R}$. La même application f est toujours linéaire, bien sûr, quand on la voit comme une fonction $\mathbb{R}^2 \to \mathbb{R}^2$. On l'appelle *la rotation d'angle* θ, ce qui doit correspondre à l'idée de rotation que vous avez étudiée au collège ou au lycée.

Exemple 15.4 – Voyons un exemple plus abstrait. Prenons

$$E = \{\, \phi\colon \mathbb{R} \to \mathbb{R} \text{ dérivable} \,\},$$

l'espace vectoriel des fonctions dérivables, et

$$F = \mathcal{F}(\mathbb{R}, \mathbb{R}),$$

l'espace vectoriel de toutes les fonctions $\mathbb{R} \to \mathbb{R}$. Alors on peut définir une application $f\colon E \to F$ par $f(\phi) = \phi'$. Cette application f est linéaire, car $(\phi+\psi)' = \phi' + \psi'$ et $(\lambda\phi)' = \lambda\phi'$ pour toute constante λ (cf proposition 6.5).

Définition 15.5 (et Proposition) – Soit $f\colon E \to F$ linéaire. On définit son *noyau* $\ker(f)$ comme étant

kernel en anglais = noyau

$$\ker(f) = \{v \in E \mid f(v) = 0\}.$$

On définit *l'image* de f, notée $\mathrm{Im}(f)$, par

$$\mathrm{Im}(f) = \{w \in F \mid \text{ il existe } v \in E \text{ tel que } w = f(v)\}.$$

On utilise aussi la notation $f(E)$ pour $\mathrm{Im}(f)$.

Alors $\ker(f)$ et $\mathrm{Im}(f)$ sont des sous-espaces vectoriels (de E et F respectivement).

La dimension de $\mathrm{Im}(f)$ est appelée le *rang* de f.

La vérification que $\ker(f)$ et $\mathrm{Im}(f)$ sont bien des sous-espaces vectoriels vous est laissée.

Exemple 15.6 – Prenons $f(v) = Av$ comme dans l'exemple 15.2. Alors $\ker(f)$ est l'ensemble des v tels que $Av = 0$: c'est l'ensemble des solutions d'un système linéaire. Réciproquement d'ailleurs, étant donné un système, on peut considérer sa matrice et l'application linéaire correspondante, dont le noyau est l'ensemble des solutions.

Avant de regarder $\mathrm{Im}(f)$ pour le même f, notons un résultat simple :

Lemme 15.7 – *Soit $f\colon E \to F$ linéaire. Si $E = \mathcal{V}ect(e_1, \ldots, e_n)$, alors $\mathrm{Im}(f) = \mathcal{V}ect(f(e_1), \ldots, f(e_n))$.*

Démonstration. Chaque $f(e_i)$ est dans $\mathrm{Im}(f)$ par définition, donc $\mathcal{V}ect(f(e_1), \ldots, f(e_n)) \subset \mathrm{Im}(f)$. Réciproquement, si $w = f(v)$, alors on écrit $v = \lambda_1 e_1 + \cdots + \lambda_n e_n$, ce qui est possible par hypothèse, et on applique f :

$$w = f(v) = f(\lambda_1 e_1 + \cdots + \lambda_n e_n) = \lambda_1 f(e_1) + \cdots + \lambda_n f(e_n).$$

Ceci montre bien que $w \in \mathcal{V}ect(f(e_1), \ldots, f(e_n))$. □

Exemple 15.8 – Reprenons encore $f\colon \mathbb{K}^n \to \mathbb{K}^m$ définie par $f(v) = Av$ comme dans l'exemple précédent, et intéressons-nous à $\mathrm{Im}(f)$.

On peut prendre la base canonique $e_1, \ldots, e_n$ de $\mathbb{K}^n$, et par le lemme on sait que $\mathrm{Im}(f) = \mathcal{V}ect(f(e_1), \ldots, f(e_n))$. Or on a

$$f(e_i) = Ae_i = A\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \text{la } i\text{-ème colonne de A.}$$

Donc $\mathrm{Im}(f)$ est l'espace engendré par les colonnes de A, c'est-à-dire que $\mathrm{Im}(f) = \mathcal{V}ect(A)$. En particulier, par définition même le rang de f coïncide avec le rang de A (rappelez-vous la définition 14.40).

Ces deux derniers exemples montrent que les deux grands types de sous-espaces de $\mathbb{K}^n$ qui nous sont familiers, à savoir ceux définis par des équations et ceux donnés comme des vects, peuvent être vus comme des noyaux ou des images d'applications linéaires. Comprendre les applications linéaires permet donc de comprendre bien des choses.

Exemple 15.9 – Reprenons l'application $f(\phi) = \phi'$ comme dans l'exemple 15.4. Le noyau de f est constitué des fonctions ϕ telles que $\phi' = 0$; d'après le théorème des accroissements finis, ceci revient à dire que ϕ est constante.

L'image de f est l'ensemble des fonctions ψ qui sont de la forme $\psi = \phi'$. En d'autres termes il s'agit des fonctions qui possèdent une primitive. Peut-on décrire facilement cet espace vectoriel ? C'est une question très difficile ! Dans le chapitre suivant nous montrerons au moins que toutes les fonctions continues possèdent une primitive (mais ça n'est qu'une description partielle de $\mathrm{Im}(f)$ bien sûr).

voir les exercices 929, 5188

Nous souhaitons décrire deux types d'applications linéaires très courantes et de nature géométrique, les projections et les symétries. Ce sont des généralisations des projections et symétries « orthogonales » que vous aviez vues au collège. Pour préparer correctement la version la plus générale, il nous faut examiner un peu les relations qu'il peut y avoir entre deux sous-espaces d'un espace donné.

On part donc d'un espace vectoriel E, et on prend deux sous-espaces U et V. On peut tout d'abord considérer l'intersection $U \cap V$, constituée des vecteurs qui sont à la fois dans U et dans V : vous vérifierez sans peine que c'est encore un sous-espace vectoriel. Une autre opération possible est la suivante.

Définition 15.10 – La *somme* des sous-espaces U et V de E, notée $U + V$, est l'ensemble des vecteurs de E de la forme $u + v$ avec $u \in U$ et $v \in V$.

Là encore, c'est un sous-espace vectoriel de E.

Exemple 15.11 – Lorsque U et V sont donnés par des équations, décrire $U \cap V$ est facile. Par exemple dans $E = \mathbb{R}^3$, si U est décrit par les équations

$$\begin{cases} 3x - y + z = 0 \\ x + 3y + 5z = 0 \end{cases}$$

et si V est décrit par l'équation $x - z = 0$, alors $U \cap V$ est l'ensemble des vecteurs dont les coordonnées vérifient toutes ces équations à la fois. En clair $U \cap V$ est décrit par

$$\begin{cases} 3x - y + z = 0 \\ x + 3y + 5z = 0 \\ x - z = 0 \end{cases}.$$

Si maintenant U et V sont donnés comme des vects, c'est $U + V$ qui est facile à décrire. En effet, les définitions entraînent immédiatement que

$$\mathcal{V}ect(u_1, \dots, u_n) + \mathcal{V}ect(v_1, \dots, v_m) = \mathcal{V}ect(u_1, \dots, u_n, v_1, \dots, v_m).$$

(Vérifiez-le.)

Si maintenant on souhaite décrire $U \cap V$ pour U et V donnés comme des vects, ou $U + V$ pour U et V donnés par des équations, la seule solution est de faire d'abord une traduction des équations aux vects ou vice-versa, comme on sait le faire.

Il existe une relation simple entre les dimensions de $U \cap V$ et $U + V$:

Proposition 15.12 – *Soient* E *un espace vectoriel et* U, V *deux sous-espaces de dimension finie. Alors* $U + V$ *et* $U \cap V$ *sont de dimension finie, et on a*

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V).$$

Démonstration. Puisque $U \cap V$ est un sous-espace de U, il est de dimension finie par le corollaire 14.39. Prenons donc une base de $U \cap V$, disons $e_1, e_2, \dots, e_d$.

D'après le théorème de la base incomplète (14.37), on peut trouver $u_1, u_2, \dots, u_k \in U$ tels que $\mathscr{B}_U = e_1, \dots, e_d, u_1, \dots, u_k$ est une base de U. De même, on peut trouver $v_1, \dots, v_\ell \in V$ tels que $\mathscr{B}_V = e_1, \dots, e_d, v_1, \dots, v_\ell$ est une base de V. Montrons que

$$\mathscr{B} = e_1, \dots, e_d, u_1, \dots, u_k, v_1, \dots, v_\ell$$

est une base de $U + V$. Ceci montrera que $U + V$ est de dimension finie, et que sa dimension est $d + k + \ell = (d + k) + (d + \ell) - d = \dim(U) + \dim(V) - \dim(U \cap V)$, comme prévu.

Pour commencer, puisque $U = \mathcal{V}ect(\mathscr{B}_U)$ et $V = \mathcal{V}ect(\mathscr{B}_V)$, il est clair que $U + V = \mathcal{V}ect(\mathscr{B})$ (voir l'exemple précédent). Donc $\mathscr{B}$ est génératrice. Pour montrer qu'elle est libre, nous devons étudier l'équation

$$\alpha_1 e_1 + \dots + \alpha_d e_d + \beta_1 u_1 + \dots + \beta_k u_k + \gamma_1 v_1 + \dots + \gamma_\ell v_\ell = 0,$$

que nous réécrivons

$$\alpha_1 e_1 + \dots + \alpha_d e_d + \beta_1 u_1 + \dots + \beta_k u_k = -(\gamma_1 v_1 + \dots + \gamma_\ell v_\ell).$$

Le membre de gauche appartient à U et le membre de droite appartient à V ; pour qu'ils soient égaux, il faut donc qu'ils appartiennent tous les deux à $U \cap V$. L'espace $U \cap V$ ayant pour base $e_1, \dots, e_d$, les deux membres de la dernière équation doivent donc être de la forme $\lambda_1 e_1 + \dots + \lambda_d e_d$ pour certains scalaires λ_i. Écrivons en particulier

$$-(\gamma_1 v_1 + \dots + \gamma_\ell v_\ell) = \lambda_1 e_1 + \dots + \lambda_d e_d,$$

ou encore

$$\lambda_1 e_1 + \dots + \lambda_d e_d + \gamma_1 v_1 + \dots + \gamma_\ell v_\ell = 0.$$

La famille $\mathscr{B}_V$ étant libre, tous les coefficients ci-dessus sont nuls : $\lambda_1 = \dots = \lambda_d = \gamma_1 = \cdots \gamma_\ell = 0$. Si nous revenons à l'équation de départ, il ne reste plus que

$$\alpha_1 e_1 + \dots + \alpha_d e_d + \beta_1 u_1 + \dots + \beta_k u_k = 0.$$

Et finalement, la famille $\mathscr{B}_U$ étant libre, ces derniers coefficients sont également nuls : $\alpha_1 = \dots = \alpha_d = \beta_1 = \dots = \beta_k = 0$. La famille $\mathscr{B}$ est bien libre. □

Corollaire 15.13 – *Soit* E *un espace vectoriel de dimension finie, et soient* U, V *deux sous-espaces. Alors, lorsque deux des trois conditions ci-dessous sont remplies, la troisième l'est également :*

1. $U \cap V = \{0\}$,
2. $E = U + V$,
3. $\dim(U) + \dim(V) = \dim(E)$.

Démonstration. Il faut simplement se rappeler les choses suivantes : si F est un sous-espace de E, alors $F = E \iff \dim(F) = \dim(E)$ (corollaire 14.39) ; en outre $F = \{0\} \iff \dim(F) = 0$. Donc on peut réécrire les trois conditions de la façon suivante :

1. $\dim(U \cap V) = 0$,
2. $\dim(E) = \dim(U) + \dim(V) - \dim(U \cap V)$,
3. $\dim(U) + \dim(V) = \dim(E)$.

(On a utilisé la proposition pour le (2).) Il est maintenant clair que si deux égalités sont vraies, alors la troisième aussi. □

Définition 15.14 – On dit que E est la *somme directe* de U et V, et on écrit $E = U \oplus V$, lorsque l'on a $U \cap V = \{0\}$ et $E = U + V$.

On dit parfois aussi que U et V sont *supplémentaires* dans E (ou encore que V est *un* supplémentaire de U dans E). Attention à ne jamais dire « complémentaire » ; attention aussi à ne pas parler « du » supplémentaire comme s'il était unique, ce n'est pas le cas.

Le corollaire indique donc que, dans le cas de la dimension finie qui est celui que nous rencontrons presque toujours, on peut vérifier si $E = U \oplus V$ de plusieurs façons. Typiquement, vérifier si $E = U + V$ peut être plus difficile que de vérifier les deux autres conditions du corollaire.

Exemple 15.15 – Prenons $E = \mathbb{R}^3$, puis U défini par l'équation $2x - y + 7z = 0$, et enfin $V = \mathcal{V}ect(v)$ avec

$$v = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

Alors $\dim(U) = 2$ (on peut prendre y et z comme paramètres), et $\dim(V) = 1$, donc $\dim(U) + \dim(V) = \dim(E)$. D'après le corollaire, pour vérifier que $E = U \oplus V$ il suffit de montrer que $U \cap V = \{0\}$. Ceci nous évite de montrer directement $E = U + V$, ce qui est un peu plus pénible.

Un vecteur de V est de la forme

$$\lambda v = \begin{pmatrix} \lambda \\ \lambda \\ 2\lambda \end{pmatrix},$$

avec $\lambda \in \mathbb{R}$. Ce vecteur est dans U lorsque

$$2\lambda - \lambda + 7 \times 2\lambda = 15\lambda = 0.$$

Ceci n'arrive que pour $\lambda = 0$, donc le seul vecteur à la fois dans U et dans V est le vecteur nul. On a bien $U \cap V = \{0\}$, et finalement $E = U \oplus V$.

Exemple 15.16 – Une situation très simple est celle d'un espace vectoriel E muni d'une base $e_1, \dots, e_n$, dans lequel on choisit de « couper en deux » ces vecteurs, en posant $U = \mathcal{V}ect(e_1, \dots, e_k)$ et $V = \mathcal{V}ect(e_{k+1}, \dots, e_n)$. Dans ce cas il est clair que $E = U \oplus V$, les deux conditions les plus faciles à vérifier étant $E = U + V$ et $\dim(U) + \dim(V) = \dim(E)$.

Il y a même une sorte de réciproque. Si $E = U \oplus V$, prenons une base $u_1, \dots, u_k$ de U et une base $v_1, \dots, v_\ell$ de V, et considérons $\mathscr{B} = u_1, \dots, u_k, v_1, \dots, v_\ell$. Alors $\mathscr{B}$ est une base de E, le plus facile étant de repérer que c'est une famille génératrice ayant $k + \ell = \dim(E)$ éléments.

voir les exercices 6871, 920, 923

La meilleure façon de comprendre de manière intuitive et géométrique les sommes directes reste d'étudier les *projections*, ce que nous allons faire maintenant.

De même que les bases nous permettent de prendre des coordonnées, les sommes directes vont nous permettre de décomposer les vecteurs. En effet, supposons que $E = U \oplus V$, et prenons $x \in E$. Puisque $E = U + V$, on peut trouver $u \in U$ et $v \in V$ tels que $x = u + v$. Mais de plus, u et v sont *uniques*. Pour vérifier ceci, écrivons $x = u' + v'$ avec $u' \in U$ et $v' \in V$, puis

$$u + v = u' + v' \Longrightarrow u - u' = v' - v .$$

On a $u - u' \in U$ et $v' - v \in V$, donc pour que ces vecteurs soient égaux, il faut qu'ils soient dans $U \cap V = \{0\}$. Ainsi $u - u' = 0 = v' - v$ et donc $u = u'$, $v = v'$.

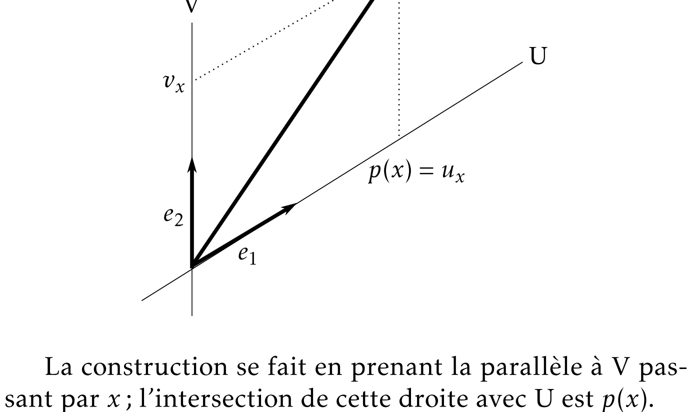
Les vecteurs u et v sont bien définis par x ; on pourrait les noter u_x et v_x. Nous allons pouvoir étudier la fonction qui à x associe u_x :

Définition 15.17 – Supposons que $E = U \oplus V$. La *projection* sur U, parallèlement à V, est l'application

$$p\colon E \longrightarrow E$$

définie par $p(x) = u_x$, où $x = u_x + v_x$ (comme ci-dessus).

Exemple 15.18 – Prenons $E = \mathbb{R}^2$, et choisissons une base e_1, e_2. Enfin, posons $U = \mathcal{V}ect(e_1)$ et $V = \mathcal{V}ect(e_2)$, de sorte que $E = U \oplus V$ comme dans l'exemple 15.16. Soit p la projection sur U, parallèlement à V. La situation se présente comme sur le dessin suivant ; on a indiqué x et $p(x)$ pour un choix de x arbitraire.



La construction se fait en prenant la parallèle à V passant par x ; l'intersection de cette droite avec U est $p(x)$.

Proposition 15.19 – *Soit p comme ci-dessus. Alors*

1. *p est linéaire,*
2. *$p(p(x)) = p(x)$,*
3. *le noyau de p est V, l'image de p est U.*

Réciproquement, si p est une application de E vers lui-même vérifiant (1) et (2), alors on a

$$E = \mathrm{Im}(p) \oplus \ker(p),$$

et p est la projection sur $\mathrm{Im}(p)$ parallèlement à $\ker(p)$.

Sur la figure précédente, essayez de voir géométriquement pourquoi $V = \ker(p)$ et $U = \mathrm{Im}(p)$.

Démonstration. Si $x = u + v$ et $y = u' + v'$, alors $x + y = (u + u') + (v + v')$ avec $u + u' \in U$ et $v + v' \in V$, donc par définition $p(x + y) = u + u'$, avec $u = p(x)$ et $u' = p(y)$. On montre de même que $p(\lambda x) = \lambda p(x)$. La projection p est bien linéaire.

En écrivant $x = u + v$ comme ci-dessus, on a $p(x) = u \in U$. Si l'on écrit $u = u + 0$, alors c'est la décomposition de u sur U et V et par définition $p(u) = u$. Ainsi $p(p(x)) = p(x)$. Les points (1) et (2) sont montrés.

Passons au (3). Si $x = u + v$ et si $p(x) = u = 0$, alors $x = v \in V$; donc $\ker(p) \subset V$. Réciproquement si $x \in V$, on écrit $x = 0 + x = u + v$ pour constater que $u = 0 = p(x)$ (et $v = x$), donc $x \in \ker(p)$ et finalement $\ker(p) = V$. L'image de p est clairement contenue dans U, et réciproquement si on prend $u \in U$, en l'écrivant $u = u + 0$ on voit (comme ci-dessus) que $p(u) = u$ donc $u \in \mathrm{Im}(p)$. L'image de p est bien U.

Voyons la réciproque, et supposons que p est linéaire de E vers E, avec $p(p(x)) = p(x)$. Si on prend $x \in E$ quelconque, on écrit simplement

$$x = p(x) + (x - p(x)). \qquad (*)$$

Bien sûr $p(x) \in \mathrm{Im}(p)$, et comme $p(x - p(x)) = p(x) - p(p(x)) = p(x) - p(x) = 0$, on a $x - p(x) \in \ker(p)$. Ceci montre que $E = \mathrm{Im}(p) + \ker(p)$.

Pour établir que la somme est directe, prenons $x \in \ker(p) \cap \mathrm{Im}(p)$. Alors $p(x) = 0$ puisque $x \in \ker(p)$. D'autre part $x = p(y)$ pour un certain y, donc $p(x) = p(p(y)) = p(y) = x$. En comparant les deux on voit que $x = 0$, et $\ker(p) \cap \mathrm{Im}(p) = \{0\}$. La somme est bien directe.

Enfin l'équation (*) montre bien que la projection de x sur $\mathrm{Im}(p)$ parallèlement à $\ker(p)$ est $p(x)$. □

Après les projections, les symétries. Cette fois-ci, au lieu de remplacer v_x par 0, on le remplace par $-v_x$. Plus précisément :

Définition 15.20 – Supposons que $E = U \oplus V$. La *symétrie* par rapport à U, dans la direction V, est l'application

$$s\colon E \longrightarrow E,$$

définie par $s(x) = u_x - v_x$, où $x = u_x + v_x$ (comme ci-dessus).

Proposition 15.21 – *Soit s comme ci-dessus. Alors*

1. *s est linéaire,*
2. *$s(s(x)) = x$,*
3. *on peut caractériser U et V par*

$$U = \{x \in E \mid s(x) = x\},$$

et

$$V = \{x \in E \mid s(x) = -x\}.$$

Réciproquement, si s est une application de E vers lui-même vérifiant (1) et (2), et si on définit U et V par les égalités ci-dessus, alors $E = U \oplus V$ et s est la symétrie par rapport à U, dans la direction V.

Démonstration. On vous laisse montrer les trois points, à titre d'exercice. Montrons la réciproque : on prend s linéaire telle que $s(s(x)) = x$ et on définit U et V par les égalités proposées. Écrivons

$$x = \frac{1}{2}(x + s(x)) + \frac{1}{2}(x - s(x)).$$

En posant $u_x = \frac{1}{2}(x + s(x))$ et $v_x = \frac{1}{2}(x - s(x))$, on a donc $x = u_x + v_x$. De plus

$$s(u_x) = \frac{1}{2}(s(x) + s(s(x))) = \frac{1}{2}(s(x) + x) = u_x,$$

donc $u_x \in U$ par définition. Un calcul similaire donne $s(v_x) = -v_x$, soit $v_x \in V$. Ceci montre que $E = U + V$.

Si $x \in U \cap V$, alors $s(x) = x = -x$, donc $2x = 0$ et $x = 0$. Par suite $U \cap V = \{0\}$, et $E = U \oplus V$. voir les exercices 5565, 970, 974

Enfin $s(x) = s(u_x + v_x) = u_x - v_x$, donc s est bien la symétrie annoncée. □

Nous avons vu qu'une matrice $A \in M_{m,n}(\mathbb{K})$ définissait une application linéaire f de $\mathbb{K}^n$ vers $\mathbb{K}^m$ par $f(v) = Av$. Nous allons voir maintenant que, réciproquement, si $f : E \to F$ est linéaire, on peut lui associer une matrice *une fois que des bases ont été choisies pour* E *et* F.

Supposons donc que $\mathscr{B} = e_1, \ldots, e_n$ est une base de E, et que $\mathscr{C} = \varepsilon_1, \ldots, \varepsilon_m$ est une base de F. Le lemme suivant contient une observation simple : pour définir une application linéaire dans cette situation, il suffit de spécifier chaque $f(e_i)$.

Lemme 15.22 – *Soient $v_1, \ldots, v_n$ des vecteurs quelconques de* F*. Alors il existe une application linéaire $f : E \to F$ et une seule telle que $f(e_i) = v_i$.*

Démonstration. Voyons l'unicité d'abord. Prenons $u \in E$, que l'on peut écrire $u = \lambda_1 e_1 + \cdots + \lambda_n e_n$. On doit avoir

$$f(u) = \lambda_1 f(e_1) + \cdots + \lambda_n f(e_n) = \lambda_1 v_1 + \cdots + \lambda_n v_n. \qquad (*)$$

Il n'y a donc qu'un seul choix possible pour $f(u)$, et f est unique si elle existe.

Pour vérifier l'existence, on définit f par la formule $(*)$, et on doit simplement vérifier qu'elle est linéaire. C'est très facile, et on vous le confie à titre d'exercice. □

Exploitons maintenant le fait que nous avons aussi choisi une base $\mathscr{C}$ pour F. En effet, pour spécifier $f(e_i)$, il suffit désormais de donner ses coordonnées dans $\mathscr{C}$; c'est-à-dire que l'on peut écrire

$$f(e_i) = a_{1i}\varepsilon_1 + a_{2i}\varepsilon_2 + \cdots + a_{mi}\varepsilon_m,$$

et l'application f est déterminée par les nombres a_{ij}. C'est la matrice (a_{ij}) que l'on appelle la matrice de f, par rapport aux bases $\mathscr{B}$ et $\mathscr{C}$.

En d'autres termes :

Définition 15.23 – Soit $f : E \to F$ une application linéaire, soit $\mathscr{B} = e_1, \ldots, e_n$ une base de E, et soit $\mathscr{C}$ une base de F. Alors la *matrice de f dans les bases $\mathscr{B}$ et $\mathscr{C}$*, que l'on va noter

$$_{\mathscr{C}}[f]^{\mathscr{B}}$$

est construite de la manière suivante : dans la colonne i, on place les coordonnées de $f(e_i)$ dans la base $\mathscr{C}$.

Lorsque les bases sont évidentes d'après le contexte, on va écrire $[f]$ plutôt que $_{\mathscr{C}}[f]^{\mathscr{B}}$.

Exemple 15.24 – Prenons $E = F = \mathbb{R}^2$, et

$$e_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

Alors $\mathscr{B} = e_1, e_2$ est une base de $\mathbb{R}^2$. On pose $U = \mathcal{V}ect(e_1)$ et $V = \mathcal{V}ect(e_2)$, de sorte que $\mathbb{R}^2 = U \oplus V$. Soit maintenant p la projection sur U, parallèlement à V.

Il y a au moins deux questions naturelles que l'on peut poser. Tout d'abord, quelle est

$$_{\mathscr{B}}[p]^{\mathscr{B}} \quad ?$$

(On dira « la matrice de p dans la base $\mathscr{B}$ » pour indiquer que l'on prend $\mathscr{C} = \mathscr{B}$.) Pour cela, on revient à la définition. Dans la colonne 1, on écrit les coordonnées de $p(e_1)$ dans la base $\mathscr{B}$. On a $p(e_1) = e_1 = 1 \times e_1 + 0 \times e_2$, donc la première colonne est

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Dans la colonne 2, on indique les coordonnées de $p(e_2)$. Or $p(e_2) = 0 = 0 \times e_1 + 0 \times e_2$, donc la deuxième colonne est

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Finalement

$$_{\mathscr{B}}[p]^{\mathscr{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Maintenant, avec $\mathbb{R}^2$ il est naturel de penser à la base canonique $\mathscr{C} = \varepsilon_1, \varepsilon_2$ avec

$$\varepsilon_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \varepsilon_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Quelle est donc

$$_{\mathscr{C}}[p]^{\mathscr{C}} \quad ?$$

Il faut calculer $p(\varepsilon_1)$ et $p(\varepsilon_2)$, dans la base canonique, et ceci nous donnera les deux colonnes de la matrice que l'on cherche. Prenons donc un vecteur quelconque

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

On va calculer ses coordonnées dans la base $\mathscr{B}$, puisque c'est avec cette base que l'on sait bien faire des calculs avec p. Pour cela introduisons

$$A = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix},$$

la matrice dont les colonnes sont les vecteurs de $\mathscr{B}$. On a alors

$$_{\mathscr{B}}[x] = A^{-1}x = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{2}{7}x_1 + \frac{1}{7}x_2 \\ -\frac{1}{7}x_1 + \frac{3}{7}x_2 \end{pmatrix}.$$

(Revoir l'exemple 14.27 si ce n'est pas clair.) En particulier on obtient donc

$$_{\mathscr{B}}[\varepsilon_1] = \begin{pmatrix} \frac{2}{7} \\ -\frac{1}{7} \end{pmatrix}, \qquad _{\mathscr{B}}[\varepsilon_2] = \begin{pmatrix} \frac{1}{7} \\ \frac{3}{7} \end{pmatrix}.$$

On a donc $\varepsilon_1 = \frac{2}{7}e_1 - \frac{1}{7}e_2$, ce qui permet de calculer que

$$p(\varepsilon_1) = \frac{2}{7}e_1 = \begin{pmatrix} \frac{6}{7} \\ \frac{2}{7} \end{pmatrix} = \frac{6}{7}\varepsilon_1 + \frac{2}{7}\varepsilon_2.$$

De la même manière, on a $\varepsilon_2 = \frac{1}{7}e_1 + \frac{3}{7}e_2$ donc

$$p(\varepsilon_2) = \frac{1}{7}e_1 = \begin{pmatrix} \frac{3}{7} \\ \frac{1}{7} \end{pmatrix} = \frac{3}{7}\varepsilon_1 + \frac{1}{7}\varepsilon_2.$$

Finalement

$$_{\mathscr{C}}[p]^{\mathscr{C}} = \begin{pmatrix} \frac{6}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{1}{7} \end{pmatrix}.$$

Pour vérifier si vous avez assimilé les définitions, assurez-vous que vous comprenez maintenant pourquoi

$$_{\mathscr{C}}[p]^{\mathscr{B}} = \begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{et} \quad _{\mathscr{B}}[p]^{\mathscr{C}} = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ 0 & 0 \end{pmatrix}.$$

Si vous trouvez cet exemple difficile à suivre, alors vous apprécierez la « formule du changement de base » que nous allons présenter très bientôt. C'est une formule simple et systématique pour organiser les calculs ci-dessus, mais il faudra toujours être capable d'écrire au moins une matrice (pour certaines bases), la formule donnera les matrices dans les autres bases. Dans l'exemple, le plus simple est de commencer par

$$_{\mathscr{B}}[p]^{\mathscr{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Nous commençons d'ailleurs avec cet exemple à voir l'intérêt de toutes ces matrices : puisqu'elles contiennent toutes la même information, à savoir une description de l'application p, libre à nous de choisir la plus simple. Et vous voyez bien que $_{\mathscr{B}}[p]^{\mathscr{B}}$ est beaucoup plus simple que les autres ! Cette idée sera « poussée » dans le chapitre « Diagonalisation ».

Commençons par constater, avec la proposition suivante, que toutes les applications linéaires (entre espaces vectoriels de dimension finie) se ramènent à multiplier une matrice par un vecteur-colonne ; et la composition des applications se ramène à multiplier les matrices.

Proposition 15.25 – *Soient* E, F *et* G *des espaces vectoriels, et soient $\mathscr{B}$, $\mathscr{C}$ et $\mathscr{D}$ des bases de ces espaces respectifs.*

1. *Si $f : E \to F$ est linéaire, et si $x \in E$, alors $[f(x)] = [f][x]$. Plus précisément*
$$_{\mathscr{C}}[f(x)] = {_{\mathscr{C}}[f]^{\mathscr{B}}} {_{\mathscr{B}}[x]}.$$
2. *Si $g : F \to G$ est linéaire, alors $[g \circ f] = [g][f]$. Plus précisément*
$$_{\mathscr{D}}[g \circ f]^{\mathscr{B}} = {_{\mathscr{D}}[g]^{\mathscr{C}}} {_{\mathscr{C}}[f]^{\mathscr{B}}}.$$

Les détails de la démonstration, qui consiste seulement à vérifier les définitions, vous sont laissés à titre d'exercice. Prenez le temps de le faire, et de vous convaincre que *c'est précisément pour avoir ce résultat que l'on a défini la multiplication des matrices comme on l'a fait*.

Exemple 15.26 – Reprenons l'exemple précédent. Nous avons calculé

$$M = {_{\mathscr{C}}[p]^{\mathscr{C}}} = \begin{pmatrix} \frac{6}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{1}{7} \end{pmatrix}.$$

Ceci nous permet de calculer l'image d'un vecteur quelconque par l'application p (dans la base canonique). En effet si

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = {_{\mathscr{C}}[x]},$$

alors

$$p(x) = {_{\mathscr{C}}[p(x)]} = {_{\mathscr{C}}[p]^{\mathscr{C}}} {_{\mathscr{C}}[x]} = Mx = \begin{pmatrix} \frac{6}{7}x_1 + \frac{3}{7}x_2 \\ \frac{2}{7}x_1 + \frac{1}{7}x_2 \end{pmatrix}.$$

En d'autres termes, l'application p n'est autre que $p(x) = Mx$. Au départ il n'était pas clair à partir de la définition « géométrique » de la projection que l'on puisse l'écrire simplement avec une matrice.

voir les exercices 2740, 2741

Formule du changement de base

Nous allons voir comment passer de la matrice d'une application linéaire, écrite dans certaines bases, à la matrice de la même application écrite dans d'autres bases. Notre point de départ est donc celui d'un espace vectoriel E muni de *deux* bases, disons $\mathscr{B} = e_1, \ldots, e_n$ et $\mathscr{C} = \varepsilon_1, \ldots, \varepsilon_n$.

Définition 15.27 – La matrice de passage de $\mathscr{B}$ à $\mathscr{C}$, que l'on va noter

$$ {}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}, $$

est la matrice obtenue de la manière suivante : dans la colonne i, on place les coordonnées de e_i dans la base $\mathscr{C}$.

(On a signalé à l'auteur que dans certains livres, ce que nous notons ${}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}$ est appelé « matrice de passage de $\mathscr{C}$ à $\mathscr{B}$ », donc en inversant les rôles de $\mathscr{B}$ et $\mathscr{C}$. Méfiance donc en lisant un énoncé d'exercice de provenance suspecte. L'important, bien sûr, est d'associer correctement une notation avec une définition, donc ${}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}$ avec la définition ci-dessus par exemple.)

Exemple 15.28 – Reprenons l'exemple 15.24, et conservons les notations. On a alors

$$ {}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}} = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix}. $$

En effet cette matrice contient bien dans la colonne 1 le vecteur e_1 de $\mathscr{B}$, écrit dans la base canonique $\mathscr{C}$, et de même la colonne 2 contient e_2.

Pour calculer ${}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}}$, on doit placer dans la colonne 1 le vecteur-colonne ${}_{\mathscr{B}}[\varepsilon_1]$, et dans la colonne 2 on place ${}_{\mathscr{B}}[\varepsilon_2]$. Nous avons fait ces calculs dans l'exemple, et finalement on a

$$ {}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}} = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix}. $$

Proposition 15.29 – *Les matrices de passage ont les propriétés suivantes :*

1. ${}_{\mathscr{B}}\mathrm{P}^{\mathscr{B}} = \mathrm{Id}$ *(matrice identité).*
2. ${}_{\mathscr{D}}\mathrm{P}^{\mathscr{B}} = {}_{\mathscr{D}}\mathrm{P}^{\mathscr{C}}{}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}$.
3. ${}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}} = \left({}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}\right)^{-1}$.

Démonstration. Le premier point est évident d'après les définitions. Pour le deuxième, le plus simple est de noter la chose suivante : si $\iota\colon \mathrm{E} \to \mathrm{E}$ est l'application $\iota(x) = x$, alors les définitions entraînent que

$$ {}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}} = {}_{\mathscr{C}}[\iota]^{\mathscr{B}}. $$

On exploite ensuite le fait que $\iota(\iota(x)) = x$ donc $\iota \circ \iota = \iota$ et par suite, en utilisant la proposition 15.25 :

$$ {}_{\mathscr{D}}\mathrm{P}^{\mathscr{B}} = {}_{\mathscr{D}}[\iota]^{\mathscr{B}} = {}_{\mathscr{D}}[\iota \circ \iota]^{\mathscr{B}} = {}_{\mathscr{D}}[\iota]^{\mathscr{C}}{}_{\mathscr{C}}[\iota]^{\mathscr{B}} = {}_{\mathscr{D}}\mathrm{P}^{\mathscr{C}}{}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}. $$

Le troisième point est maintenant facile, puisque l'on a

$$ {}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}}{}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}} = {}_{\mathscr{B}}\mathrm{P}^{\mathscr{B}} = \mathrm{Id}, $$

donc les matrices ${}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}$ et ${}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}}$ sont bien inverses l'une de l'autre, comme annoncé. □

Voici maintenant la formule proprement dite :

Proposition 15.30 – *Soit $f\colon \mathrm{E} \to \mathrm{F}$ linéaire, soient $\mathscr{B}$ et $\mathscr{B}'$ deux bases de E, et soient $\mathscr{C}$ et $\mathscr{C}'$ deux bases de F. Alors*

$$ {}_{\mathscr{C}'}[f]^{\mathscr{B}'} = {}_{\mathscr{C}'}\mathrm{P}^{\mathscr{C}}{}_{\mathscr{C}}[f]^{\mathscr{B}}{}_{\mathscr{B}}\mathrm{P}^{\mathscr{B}'}. $$

Cette formule est plus facile à mémoriser qu'il n'y paraît. Il suffit de se rappeler qu'il faut « mettre » des matrices de passage à gauche et à droite ; ensuite, pour écrire les bonnes bases, il suffit de s'assurer d'abord que les bases apparaissant côte-à-côte sont les mêmes (ci-dessus, on a écrit $\mathscr{B}$ deux fois, côte-à-côte, et de même pour $\mathscr{C}$) ; enfin, les bases « à l'extérieur » des formules (donc $\mathscr{B}'$ et $\mathscr{C}'$) sont les mêmes des deux côtés de l'égalité.

Démonstration. On utilise la même astuce. Soit $\iota_{\mathrm{E}}\colon \mathrm{E} \to \mathrm{E}$ l'application $\iota_{\mathrm{E}}(x) = x$, et soit ι_{F} défini de la même manière. On a $\iota_{\mathrm{F}}(f(\iota_{\mathrm{E}}(x))) = f(x)$, donc $\iota_{\mathrm{F}} \circ f \circ \iota_{\mathrm{E}} = f$, ce qui donne en termes de matrices (en utilisant la proposition 15.25) :

$$ {}_{\mathscr{C}'}[f]^{\mathscr{B}'} = {}_{\mathscr{C}'}[\iota_{\mathrm{F}} \circ f \circ \iota_{\mathrm{E}}]^{\mathscr{B}'} = {}_{\mathscr{C}'}[\iota_{\mathrm{F}}]^{\mathscr{C}}{}_{\mathscr{C}}[f]^{\mathscr{B}}{}_{\mathscr{B}}[\iota_{\mathrm{E}}]^{\mathscr{B}'} = {}_{\mathscr{C}'}\mathrm{P}^{\mathscr{C}}{}_{\mathscr{C}}[f]^{\mathscr{B}}{}_{\mathscr{B}}\mathrm{P}^{\mathscr{B}'}, $$

puisque ${}_{\mathscr{C}'}[\iota_{\mathrm{F}}]^{\mathscr{C}} = {}_{\mathscr{C}'}\mathrm{P}^{\mathscr{C}}$ et ${}_{\mathscr{B}'}[\iota_{\mathrm{E}}]^{\mathscr{B}} = {}_{\mathscr{B}'}\mathrm{P}^{\mathscr{B}}$. □

Exemple 15.31 – Reprenons l'exemple 15.24 : les choses vont être maintenant beaucoup plus simples. Il s'agit donc de la projection $p\colon \mathbb{R}^2 \to \mathbb{R}^2$ sur $\mathrm{U} = \mathit{Vect}(e_1)$ parallèlement à $\mathrm{V} = \mathit{Vect}(e_2)$. Dans la base $\mathscr{B} = e_1, e_2$ on a

$$ {}_{\mathscr{B}}[p]^{\mathscr{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, $$

puisque $p(e_1) = e_1$ et $p(e_2) = 0$ (pour l'instant les vecteurs e_1 et e_2 particuliers que l'on choisit ne changent rien à l'affaire). Si maintenant on considère la base canonique $\mathscr{C}$ et que l'on veut la matrice de p dans cette base, on utilise la formule du changement de base :

$$ {}_{\mathscr{C}}[p]^{\mathscr{C}} = {}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}{}_{\mathscr{B}}[p]^{\mathscr{B}}{}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}}. $$

Dans l'exemple 15.28, nous avons vu que

$$ {}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}} = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix}. $$

Pour l'autre, on utilise le fait que ${}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}} = \left({}_{\mathscr{C}}\mathrm{P}^{\mathscr{B}}\right)^{-1}$. Nous avons déjà fait ce calcul, et on trouve

$$ {}_{\mathscr{B}}\mathrm{P}^{\mathscr{C}} = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix}. $$

Ce qui donne bien

$$ {}_{\mathscr{C}}[p]^{\mathscr{C}} = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix} = \begin{pmatrix} \frac{6}{7} & \frac{3}{7} \\ \frac{2}{7} & \frac{1}{7} \end{pmatrix}. $$

On peut traiter sans peine le cas de la projection s par rapport à U, dans la direction V. En effet on a $s(e_1) = e_1$ et $s(e_2) = -e_2$, donc

$$ {}_{\mathscr{B}}[s]^{\mathscr{B}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. $$

Dans la base canonique, on introduit exactement les mêmes matrices de passage, donc

$$ {}_{\mathscr{C}}[s]^{\mathscr{C}} = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix} = \begin{pmatrix} \frac{5}{7} & \frac{6}{7} \\ \frac{4}{7} & -\frac{5}{7} \end{pmatrix}. $$

Ce qui signifie que pour trouver l'image d'un vecteur quelconque par cette symétrie, on peut calculer simplement

$$ s\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{5}{7} & \frac{6}{7} \\ \frac{4}{7} & -\frac{5}{7} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{5}{7}x_1 + \frac{6}{7}x_2 \\ \frac{4}{7}x_1 - \frac{5}{7}x_2 \end{pmatrix}. $$

voir les exercices 2433, 3400, 5259

– Fin de la première lecture –

Le lecteur est invité à revoir les définitions des termes « injectif », « surjectif », et « bijectif », introduits dans le tout premier chapitre de ce livre.

Nous allons examiner ces concepts dans le cadre des applications linéaires. Il se trouve que la situation est bien plus simple que dans le cas général. Commençons par :

LEMME 15.32 – *Soit $f\colon E \to F$ linéaire. Si f possède une réciproque $f^{-1}\colon F \to E$, alors f^{-1} est également linéaire.*

Démonstration. Prenons u et v dans F, et soit $x = f^{-1}(u+v)$. On a $f(x) = u + v = f(f^{-1}(u)) + f(f^{-1}(v)) = f(f^{-1}(u) + f^{-1}(v))$ puisque f est linéaire. En appliquant f^{-1}, on obtient $x = f^{-1}(u) + f^{-1}(v) = f^{-1}(u+v)$. On vous laisse montrer de la même manière que $f^{-1}(\lambda v) = \lambda f^{-1}(v)$. □

On utilise un mot savant pour les applications bijectives et linéaires :

DÉFINITION 15.33 – Une application linéaire et bijective est appelée un *isomorphisme*. Lorsqu'il existe un isomorphisme $E \to F$, on dit que E et F sont *isomorphes*.

Ce nouveau nom ne doit pas cacher un vieux calcul :

PROPOSITION 15.34 – *Soit $f\colon E \to F$ linéaire, soit $\mathscr{B}$ une base (finie) de E, et soit $\mathscr{C}$ une base (finie) de F. Alors f est un isomorphisme si et seulement si la matrice ${}_{\mathscr{C}}[f]^{\mathscr{B}}$ est inversible. De plus la matrice de la réciproque f^{-1} est l'inverse de la matrice de f.*

Ce qui veut dire que l'on peut se ramener à un calcul de déterminant.

Démonstration. Si f^{-1} existe, on note que $f^{-1}(f(x)) = x$ donc la matrice de $f^{-1}\circ f$ dans la base $\mathscr{B}$ est l'identité. Ainsi

$$\mathrm{Id} = {}_{\mathscr{B}}\left[f^{-1}\circ f\right]^{\mathscr{B}} = {}_{\mathscr{B}}\left[f^{-1}\right]^{\mathscr{C}}{}_{\mathscr{C}}[f]^{\mathscr{B}}\,.$$

De la même manière, on montre dans l'autre sens que

$${}_{\mathscr{C}}[f]^{\mathscr{B}}{}_{\mathscr{B}}\left[f^{-1}\right]^{\mathscr{C}} = \mathrm{Id}\,,$$

ce qui montre que

$${}_{\mathscr{B}}\left[f^{-1}\right]^{\mathscr{C}} = \left({}_{\mathscr{C}}[f]^{\mathscr{B}}\right)^{-1}.$$

Réciproquement, si la matrice de f est inversible, on définit

$$M = {}_{\mathscr{C}}[f]^{\mathscr{B}} \quad \text{et} \quad N = M^{-1}\,,$$

et on écrit g pour l'unique application $F \to E$ dont la matrice est

$${}_{\mathscr{B}}[g]^{\mathscr{C}} = N.$$

Les relations $MN = NM = \mathrm{Id}$ entraînent $g(f(x)) = x$ et $f(g(x)) = x$, donc $g = f^{-1}$. □

Une matrice inversible se doit d'être carrée, donc citons tout de suite :

COROLLAIRE 15.35 – *Deux espaces de dimension finie E et F sont isomorphes $\Longleftrightarrow$ ils ont la même dimension.*

Démonstration. Pour toute application linéaire $f\colon E \to F$, sa matrice est de dimension $m\times n$, avec $n = \dim(E)$ et $m = \dim(F)$. S'il en existe une qui est bijective, alors sa matrice doit être carrée, donc $n = m$. Réciproquement si $n = m$, prenons n'importe quelle matrice A inversible de taille $n \times n$ (par exemple l'identité), prenons une base $\mathscr{B}$ de E et une base $\mathscr{C}$ de F, et enfin prenons f l'unique application linéaire $f\colon E \to F$ telle que

$${}_{\mathscr{C}}[f]^{\mathscr{B}} = A.$$

C'est un isomorphisme d'après la proposition. □

EXEMPLE 15.36 – Prenons $r\colon \mathbb{R}^2 \to \mathbb{R}^2$ la rotation d'angle θ autour de l'origine. Rappelons qu'en identifiant $\mathbb{R}^2$ avec le plan complexe $\mathbb{C}$, on a $r(z) = e^{i\theta}z$. Choisissons la base $\mathscr{B} = 1, i$. On a $r(1) = e^{i\theta} = \cos(\theta) + i\sin(\theta)$, et $r(i) = ie^{i\theta} = -\sin(\theta) + i\cos(\theta)$. Par définition on a donc

$${}_{\mathscr{B}}[r]^{\mathscr{B}} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Le déterminant de cette matrice est $\cos(\theta)^2 + \sin(\theta)^2 = 1 \neq 0$, donc elle est inversible et son inverse est

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Donc r est bijective et sa réciproque est donnée par la matrice ci-dessus dans la base $\mathscr{B}$. Puisque $\cos(-\theta) = \cos(\theta)$ et $\sin(-\theta) = -\sin(\theta)$, on peut réécrire

$$[r^{-1}] = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix}.$$

On constate que la réciproque de r est la rotation d'angle $-\theta$, ce qui est normalement évident du point de vue géométrique.

Les conditions de surjectivité et d'injectivité se vérifient également facilement. On a d'abord :

PROPOSITION 15.37 – *Une application linéaire $f\colon E \to F$ est surjective $\Longleftrightarrow$ le rang de f vaut* $\dim(F)$.

Démonstration. Tout est dans les définitions. On a f surjective $\Leftrightarrow \mathrm{Im}(f) = F \Leftrightarrow \dim(\mathrm{Im}(f)) = \dim(F)$, et bien sûr $\dim(\mathrm{Im}(f))$ est par définition le rang de f. □

EXEMPLE 15.38 – Une application linéaire $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ ne peut pas être surjective. En effet, si A désigne la matrice de f dans les bases canoniques, alors $f(v) = Av$ et le rang de f est le rang de la matrice A. Or le rang d'une matrice 3×2 ne saurait être égal à 3.

Pour l'injectivité, le résultat suivant est très utile :

PROPOSITION 15.39 – *Une application linéaire f est injective $\Longleftrightarrow$ on a* $\ker(f) = \{0\}$.

Démonstration. Si f est injective, alors pour $x \in \ker(f)$ on a $f(x) = 0 = f(0)$ donc $x = 0$, et $\ker(f) = \{0\}$. Réciproquement, supposons que $\ker(f) = \{0\}$. Si $f(x_1) = f(x_2)$, alors $f(x_1 - x_2) = f(x_1) - f(x_2) = 0$, donc $x_1 - x_2 \in \ker(f)$, d'où $x_1 = x_2$, et f est bien injective. □

EXEMPLE 15.40 – Voici la traduction en termes de matrices. Prenons $f\colon \mathbb{K}^n \to \mathbb{K}^m$ définie par $f(v) = Av$ pour une matrice A. Alors les éléments v de $\ker(f)$ sont les solutions du système linéaire $Av = 0$.

Dire que $\ker(f) = \{0\}$, c'est affirmer que ce système n'a que la solution nulle. On s'aperçoit alors que l'injectivité de f revient à exiger que *les colonnes de* A *forment une famille libre*.

En particulier, c'est impossible pour $n > m$ (théorème 14.30), et il n'y a aucune application linéaire injective $\mathbb{R}^3 \to \mathbb{R}^2$ par exemple.

voir l'exercice 956

On voit que des considérations simples sur les matrices nous permettent de faire des liens entre l'injectivité ou la surjectivité d'une application linéaire et les dimensions des espaces qui sont en jeu. En fait il y a une façon très simple et très générale de résumer toutes ces relations, qu'il faut retenir : c'est le « théorème du rang » que nous présentons maintenant.

C'est le suivant :

Théorème 15.41 – *Soit* $f\colon \mathrm{E} \to \mathrm{F}$ *une application linéaire, et supposons que* E *est de dimension finie. Alors* $\mathrm{Im}(f)$ *est de dimension finie, et on a*

$$\dim(\mathrm{E}) = \dim(\ker(f)) + \dim(\mathrm{Im}(f)).$$

Démonstration. Soit $e_1, \ldots, e_k$ une base de $\ker(f)$. Par le théorème de la base incomplète, on peut trouver $\varepsilon_1, \ldots, \varepsilon_\ell$ tels que la famille $\mathscr{B} = e_1, \ldots, e_k, \varepsilon_1, \ldots, \varepsilon_\ell$ est une base de E. Posons $f_i = f(\varepsilon_i)$, et montrons que la famille $\mathscr{C} = f_1, \ldots, f_\ell$ est une base de $\mathrm{Im}(f)$. Comme $\dim(\mathrm{E}) = k + \ell$, le théorème sera établi.

Si $y \in \mathrm{Im}(f)$, par définition $y = f(x)$ pour un certain $x \in \mathrm{E}$, et on peut écrire

$$x = \lambda_1 e_1 + \cdots + \lambda_k e_k + \mu_1 \varepsilon_1 + \cdots + \mu_\ell \varepsilon_\ell,$$

d'où

$$y = f(x) = \mu_1 f(\varepsilon_1) + \cdots + \mu_\ell f(\varepsilon_\ell) = \mu_1 f_1 + \cdots + \mu_\ell f_\ell$$

(car $f(e_i) = 0$ bien sûr). Donc $\mathscr{C}$ est génératrice de $\mathrm{Im}(f)$.

Si maintenant on a une combinaison linéaire nulle :

$$\mu_1 f_1 + \cdots + \mu_\ell f_\ell = 0 = \mu_1 f(\varepsilon_1) + \cdots + \mu_\ell f(\varepsilon_l) = f(\mu_1 \varepsilon_1 + \cdots + \mu_\ell \varepsilon_\ell),$$

on en conclut que $\mu_1 \varepsilon_1 + \cdots + \mu_\ell \varepsilon_\ell \in \ker(f)$. Nous avons une base de $\ker(f)$ à notre disposition, écrivons donc qu'il existe des scalaires $\lambda_1, \ldots, \lambda_k$ tels que

$$\mu_1 \varepsilon_1 + \cdots + \mu_\ell \varepsilon_\ell = \lambda_1 e_1 + \cdots + \lambda_k e_k.$$

On en déduit

$$\lambda_1 e_1 + \cdots + \lambda_k e_k - \mu_1 \varepsilon_1 - \cdots - \mu_\ell \varepsilon_\ell = 0,$$

et comme $\mathscr{B}$ est une base de E, on a finalement $\lambda_1 = \cdots = \lambda_k = \mu_1 = \cdots = \mu_\ell = 0$. En particulier, la famille $\mathscr{C}$ est libre. □

Exemple 15.42 – Pour une application $f\colon \mathbb{K}^n \to \mathbb{K}^m$, de la forme $f(v) = \mathrm{A}v$, où A est une matrice, ce théorème dit une chose bien concrète en termes du système $\mathrm{A}v = 0$. En effet $\dim(\mathrm{E}) = n$ est le nombre d'inconnues en jeu, $\dim(\ker(f))$ est le nombre (minimal) de paramètres nécessaires pour décrire l'ensemble des solutions, et $\dim(\mathrm{Im}(f))$ est le rang de la matrice A. On a donc

$$\left(\begin{array}{c}\text{nombre}\\ \text{d'inconnues}\end{array}\right) = \left(\begin{array}{c}\text{nombre de}\\ \text{paramètres}\end{array}\right) + \left(\begin{array}{c}\text{rang de}\\ \text{la matrice}\end{array}\right).$$

Avec un peu d'expérience vis-à-vis des systèmes, ça n'a rien de surprenant. Il est toutefois appréciable d'avoir une formulation très précise de cette égalité – par exemple la notion de dimension d'un espace vectoriel rend précise l'idée de nombre minimal de paramètres nécessaires pour décrire les solutions.

Le théorème du rang a de nombreuses conséquences immédiates.

Corollaire 15.43 – *Soit* $f\colon \mathrm{E} \to \mathrm{F}$ *linéaire. Alors*

1. *Si* f *est injective, on a* $\dim(\mathrm{E}) \leq \dim(\mathrm{F})$.
2. *Si* f *est surjective, on a* $\dim(\mathrm{E}) \geq \dim(\mathrm{F})$.
3. *Si* f *est un isomorphisme, on a* $\dim(\mathrm{E}) = \dim(\mathrm{F})$.

Nous avions observé ces phénomènes un peu plus haut, mais la démonstration est maintenant plus directe :

Démonstration. Si f est injective, alors $\dim(\ker(f)) = 0$ (proposition 15.39), d'où $\dim(\mathrm{E}) = \dim(\mathrm{Im}(f)) \leq \dim(\mathrm{F})$ puisque $\mathrm{Im}(f)$ est un sous-espace de F.

Si f est surjective, alors $\dim(\mathrm{Im}(f)) = \dim(\mathrm{F}) = \dim(\mathrm{E}) - \dim(\ker(f)) \leq \dim(\mathrm{E})$. □

Le prochain résultat met également au clair quelque chose que nous avions observé sur des exemples :

Corollaire 15.44 – *Soit* $f\colon \mathrm{E} \to \mathrm{F}$ *linéaire. Si deux des propriétés ci-dessous sont satisfaites, alors la troisième l'est également :*

1. f *est injective,*
2. f *est surjective,*
3. $\dim(\mathrm{E}) = \dim(\mathrm{F})$.

Démonstration. Dans le précédent corollaire on a vu que (1) et (2) entraînent (3). Supposons que l'on ait (1) et (3). Alors

$$\dim(\mathrm{Im}(f)) = \dim(\mathrm{E}) - \dim(\ker(f)) = \dim(\mathrm{F}) - 0 = \dim(\mathrm{F}),$$

donc $\mathrm{Im}(f) = \mathrm{F}$ et f est surjective. Si l'on a (2) et (3), alors

$$\dim(\ker(f)) = \dim(\mathrm{E}) - \dim(\mathrm{Im}(f)) = \dim(\mathrm{F}) - \dim(\mathrm{F}) = 0,$$

donc $\ker(f) = \{0\}$ et f est injective. □

voir les exercices 934, 941, 943, 954, 959

Nous verrons de nombreuses applications dans les exercices.

Le théorème du rang occupe une place centrale en algèbre linéaire. À tel point que dans certains livres sur le sujet, on trouve une démonstration de ce théorème très tôt dans l'exposition, avec les autres résultats présentés comme conséquences. Ce genre d'approche est plus concis mais plus difficile à suivre pour les débutants. Il est probable qu'en deuxième année, on vous donne un résumé de l'algèbre linéaire de première année qui soit de ce genre.

Pour se faire une idée, voici de nouvelles démonstrations de résultats déjà obtenus, qui font usage du théorème du rang. Notez la concision des arguments – en contrepartie de leur côté abstrait. Il est naturel que ces démonstrations soient plus difficiles à suivre pour l'instant.

Lemme 15.45 – *Soient* A *et* B *des matrices carrées telles que* $AB = \mathrm{Id}$. *Alors on a également* $BA = \mathrm{Id}$, *et* $B = A^{-1}$

Nous avions vu ça en tant que lemme 12.23, et la démonstration faisait appel à la notion de matrice bien échelonnée.

Démonstration. Soit $E = M_n(\mathbb{K})$, vu comme un espace vectoriel de dimension finie, et soit

$$f\colon \begin{array}{ccc} M_n(\mathbb{K}) & \longrightarrow & M_n(\mathbb{K}) \\ M & \mapsto & f(M) = BM. \end{array}$$

On voit tout de suite que f est linéaire. Montrons qu'elle est injective. Si $f(M) = 0$, alors $BM = 0$; en multipliant par A à gauche, on obtient $ABM = \mathrm{Id}\, M = M = 0$, donc $\ker(f) = \{0\}$. Ainsi f est injective, et d'après le corollaire 15.44, elle est également surjective (on a $E = F$ ici). On conclut qu'il existe, en particulier, une matrice C telle que $f(C) = \mathrm{Id}$, soit $BC = \mathrm{Id}$. En multipliant encore par A à gauche, on a $C = A$. □

Voici maintenant le résultat selon lequel le rang d'une matrice est égal au rang de sa transposée (voir le théorème 14.47), dans ce nouveau style. On obtient même un peu plus qu'avant. Quelques notations : on travaille avec des matrices $m \times n$, et on écrit

$$I_r^{m\times n} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix},$$

c'est-à-dire la matrice dont tous les coefficients sont nuls, sauf les r premiers sur la diagonale, qui valent 1. Lorsque la taille est évidente, on écrit juste I_r.

Proposition 15.46 – *Soit* A *une matrice* $m \times n$.

1. *Si* $P \in M_n(\mathbb{K})$ *est inversible, alors* $\mathrm{rang}(AP) = \mathrm{rang}(A)$.
2. *Si* $Q \in M_m(\mathbb{K})$ *est inversible, alors* $\mathrm{rang}(QA) = \mathrm{rang}(A)$.
3. $\mathrm{rang}(A) = r \iff$ *il existe deux matrices* P *et* Q *inversibles telles que* $QAP = I_r$.
4. $\mathrm{rang}(A) = \mathrm{rang}({}^tA)$.

Nous avons vu tous ces résultats, à part le (3). La démonstration va être très différente.

Démonstration. La matrice A donne une application linéaire que l'on va noter $f_A\colon \mathbb{K}^n \to \mathbb{K}^m$, définie par $f_A(v) = Av$. De même on note f_P et f_Q pour les applications linéaires correspondant à P et Q.

Montrons le (1). Le rang de A est $\dim(\mathrm{Im}(f_A)) = \dim(f_A(\mathbb{K}^n))$, et de même le rang de AP est $\dim(f_A(f_P(\mathbb{K}^n)))$. Puisque f_P est un isomorphisme, elle est surjective, et donc $f_P(\mathbb{K}^n) = \mathbb{K}^n$. D'où le (1).

Pour le (2), on note que le rang de QA est $\dim(f_Q(f_A(\mathbb{K}^n)))$. L'application $g\colon f_A(\mathbb{K}^n) \to f_Q(f_A(\mathbb{K}^n))$ donnée par $g(v) = f_Q(v)$ est surjective par définition, et injective parce que f_Q est elle-même injective (comme application définie sur $\mathbb{K}^m$ tout entier). Donc g est un isomorphisme et on en conclut que $\dim(f_A(\mathbb{K}^n)) = \dim(f_Q(f_A(\mathbb{K}^n)))$. D'où le (2).

Pour le (3), si $QAP = I_r$ alors d'après les points (1) et (2) $\mathrm{rang}(A) = \mathrm{rang}(QAP) = \mathrm{rang}(I_r)$, et bien sûr $\mathrm{rang}(I_r) = r$. Réciproquement, supposons que le rang de A est r. Comme dans la démonstration du théorème du rang, prenons $e_1, \ldots, e_k$ une base de $\ker(f_A)$, que l'on complète en une base de $\mathbb{K}^n$ avec des vecteurs $\varepsilon_1, \ldots, \varepsilon_r$; on a vu que l'on obtient une base de $\mathrm{Im}(f_A)$ en prenant $f_1, \ldots, f_r$ où $f_i = f_A(\varepsilon_i)$. Enfin, complétons $f_1, \ldots, f_r$ en une base de $\mathbb{K}^m$ en ajoutant des vecteurs $f_{r+1}, \ldots, f_m$. Si $\mathscr{B} = \varepsilon_1, \ldots, \varepsilon_r, e_1, \ldots, e_k$ et $\mathscr{C} = f_1, \ldots, f_m$, alors par définition

$${}_{\mathscr{C}}[f_A]^{\mathscr{B}} = I_r.$$

Mais alors ${}_{\mathscr{C}}[f_A]^{\mathscr{B}} = QAP$ où Q et P sont des matrices de passage bien choisies (et en particulier inversibles). Ceci donne le (3).

Le (4) est maintenant évident. En effet si A est de rang r, on a $QAP = I_r^{m\times n}$ d'où ${}^tP\,{}^tA\,{}^tQ = {}^tI_r^{m\times n} = I_r^{n\times m}$. D'après le (3) appliqué à ${}^tA$, on en déduit que ${}^tA$ est de rang r également. □ *voir l'exercice 3351*

Chapitre 16

Diagonalisation

On rappelle que la lettre $\mathbb{K}$ désigne $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$.

Dans ce chapitre, les concepts d'algèbre linéaire des chapitres précédents vont être mis en application. Nous allons décrire la technique générale de la *diagonalisation*, qui sera utilisée dans ce livre dans le cadre de problèmes très concrets : lors de l'étude de certaines équations différentielles d'une part (décrite dans le chapitre 10), et d'autre part pour analyser certaines suites récurrentes. Voyons ce dernier point tout de suite.

Imaginons une suite de vecteurs $(X_n)_{n\geq 0}$, avec $X_n \in \mathbb{R}^d$, définie par récurrence de la manière suivante : on se donne X_0, on fixe une matrice A de dimension $d\times d$, et on pose

$$X_{n+1} = AX_n. \qquad (*)$$

Pour être très concret, nous allons prendre un exemple célèbre. Commençons par la *suite de Fibonacci* $(u_n)_{n\geq 0}$, qui est la suite de réels définie par $u_0 = u_1 = 1$ et $u_{n+2} = u_{n+1} + u_n$. On peut alors poser pour tout $n \geq 0$:

$$X_n = \begin{pmatrix} u_n \\ u_{n+1} \end{pmatrix} \in \mathbb{R}^2.$$

On a alors

$$X_{n+1} = \begin{pmatrix} u_{n+1} \\ u_{n+2} \end{pmatrix} = \begin{pmatrix} u_{n+1} \\ u_{n+1}+u_n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} u_n \\ u_{n+1} \end{pmatrix} = AX_n,$$

où A est la matrice 2×2 ci-dessus. On est donc bien en présence d'une suite de la forme (*). Nous allons voir que l'on sait bien mieux étudier (X_n) que (u_n). En fait en passant par (X_n), nous allons trouver une expression directe pour u_n en fonction de n, ce qui n'est *a priori* pas évident du tout.

Commençons par quelques calculs :

$$X_1 = AX_0, \quad X_2 = AX_1 = A^2X_0, \quad X_3 = AX_2 = A^3X_0,$$

et par récurrence on obtient immédiatement $X_n = A^nX_0$. Nous devons donc calculer les puissances successives de la matrice A. Par le calcul direct, c'est difficile (essayez quelques valeurs de n pour tenter de deviner la formule).

Soit alors maintenant

$$P = \begin{pmatrix} \frac{-1+\sqrt5}{2} & \frac{-1-\sqrt5}{2} \\ 1 & 1 \end{pmatrix}.$$

D'où sort cette matrice ? Tout le but de ce chapitre, justement, est d'expliquer d'où provient P, et comment la trouver par vous-même. Pour l'instant, supposons donc que l'on ait envie d'essayer cette matrice, et de calculer $P^{-1}AP$. On trouve

$$P^{-1} = \begin{pmatrix} \frac{\sqrt5}{5} & \frac{\sqrt5}{10}+\frac12 \\ -\frac{\sqrt5}{5} & -\frac{\sqrt5}{10}+\frac12 \end{pmatrix},$$

puis

$$P^{-1}AP = \begin{pmatrix} \frac{1+\sqrt5}{2} & 0 \\ 0 & \frac{1-\sqrt5}{2} \end{pmatrix}.$$

Voilà qui nous arrange bien. En effet, la matrice $P^{-1}AP$ est *diagonale*, c'est-à-dire que seuls les coefficients sur sa diagonale sont non-nuls ; on peut donc calculer les puissances de cette matrice sans effort :

$$(P^{-1}AP)^n = \begin{pmatrix} \left(\frac{1+\sqrt5}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt5}{2}\right)^n \end{pmatrix}.$$

Bien sûr ce que nous cherchons, ce sont les puissances de A. Mais à bien y regarder on a :

$$(P^{-1}AP)^2 = P^{-1}APP^{-1}AP = P^{-1}A^2P,$$

et de même

$$(P^{-1}AP)^3 = P^{-1}AP(P^{-1}AP)^2 = P^{-1}APP^{-1}A^2P = P^{-1}A^3P.$$

Par récurrence on obtient pour tout n :

$$(P^{-1}AP)^n = P^{-1}A^nP.$$

Ainsi la matrice que l'on cherche est tout simplement

$$A^n = P(P^{-1}AP)^nP^{-1} = P\begin{pmatrix} \left(\frac{1+\sqrt5}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt5}{2}\right)^n \end{pmatrix}P^{-1}.$$

Il n'y a plus qu'à multiplier ces trois matrices. Pour le faire sans douleur, introduisons

$$\lambda_1 = \frac{1+\sqrt5}{2}, \qquad \lambda_2 = \frac{1-\sqrt5}{2},$$

de sorte que

$$P = \begin{pmatrix} -\lambda_2 & -\lambda_1 \\ 1 & 1 \end{pmatrix} \quad \text{et} \quad P^{-1} = \frac{1}{\sqrt5}\begin{pmatrix} 1 & \lambda_1 \\ -1 & -\lambda_2 \end{pmatrix}.$$

En tenant compte de la relation $\lambda_1\lambda_2 = -1$, on obtient finalement

$$A^n = \frac{1}{\sqrt5}\begin{pmatrix} \lambda_1^{n-1}-\lambda_2^{n-1} & \lambda_1^n-\lambda_2^n \\ \lambda_1^n-\lambda_2^n & \lambda_1^{n+1}-\lambda_2^{n+1} \end{pmatrix}.$$

Récoltons le fruit de nos efforts, et retournons à la suite de Fibonacci. Nous avions

$$\begin{pmatrix} u_n \\ u_{n+1} \end{pmatrix} = X_n = A^nX_0 = A^n\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Sur la première ligne de cette égalité de matrices-colonnes, on a la relation

$$u_n = \frac{1}{\sqrt5}\left(\lambda_1^{n-1}-\lambda_2^{n-1}+\lambda_1^n-\lambda_2^n\right),$$

ce qui est donc une formule donnant le n-ième nombre de Fibonacci. Il est remarquable que ce soit un nombre entier, de ce point de vue. Noter qu'en utilisant les relations $1+\lambda_1 = \lambda_1^2$ et $1+\lambda_2 = \lambda_2^2$, on peut même simplifier cette expression en

$$u_n = \frac{1}{\sqrt5}\left(\lambda_1^{n+1}-\lambda_2^{n+1}\right).$$

Pouvez-vous utiliser cette expression pour u_n afin de calculer la limite de $\frac{u_{n+1}}{u_n}$?

Le calcul ci-dessus a été rendu possible par l'arrivée dramatique de la matrice P, ayant la propriété que $P^{-1}AP$ est diagonale. Trouver P, étant donnée la matrice A, est ce qu'on appelle *diagonaliser* A. Dans ce chapitre nous allons voir comment procéder (lorsque c'est possible). Nous aurons besoin d'utiliser les concepts d'espace vectoriel, d'application linéaire, mais aussi de déterminant, développés dans les chapitres précédents. En un sens, nous voyons une mise en oeuvre de toute cette théorie.

Nous allons commencer par donner des noms aux phénomènes observés dans l'exemple précédent. Ce chapitre introduit beaucoup de vocabulaire ! Précisons que, pour cette raison, nous donnerons vers la fin du chapitre un résumé. Vous verrez qu'on en arrive finalement à une méthode vraiment systématique (qu'un ordinateur est capable de suivre, d'ailleurs). Mais nous avons du chemin à parcourir avant de pouvoir montrer ça.

DÉFINITION 16.1 – Deux matrices carrées A et B à coefficients dans $\mathbb{K}$ sont dites *conjuguées*, ou *semblables*, s'il existe une matrice inversible P à coefficients dans $\mathbb{K}$ telle que $B = P^{-1}AP$ (ou ce qui revient au même, $A = PBP^{-1}$).

On dit qu'une matrice A est *diagonalisable* lorsqu'il existe une matrice diagonale D telle que A et D sont conjuguées.

EXEMPLE 16.2 – Nous avons vu ci-dessus que la matrice

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

et la matrice

$$D = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}$$

sont conjuguées : en effet $D = P^{-1}AP$ avec

$$P = \begin{pmatrix} \frac{-1+\sqrt{5}}{2} & \frac{-1-\sqrt{5}}{2} \\ 1 & 1 \end{pmatrix}.$$

En particulier, la matrice A est diagonalisable.

En présence de deux matrices A et B, il est difficile de savoir si elles sont conjuguées, et dans ce chapitre nous allons apprendre quelques techniques. Voici déjà un premier critère simple.

LEMME 16.3 – *Si A et B sont conjuguées, alors* $\det(A) = \det(B)$.

Démonstration. En effet, si $B = P^{-1}AP$, alors

$$\det(B) = \det(P)^{-1}\det(A)\det(P) = \det(A).$$

□

EXEMPLE 16.4 – Dans l'exemple précédent, on a bien $\det(A) = \det(D) = -1$. Par contre les matrices

$$A = \begin{pmatrix} 2 & -3 \\ 4 & -5 \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

ne sont pas conjuguées, puisque $\det(A) = 2$ et $\det(B) = 0$.

Après le déterminant, voici la « trace » :

DÉFINITION 16.5 – Soit $A = (a_{ij})$ une matrice carrée. La *trace* de A, notée Tr(A), est la somme des coefficients sur la diagonale de A.

EXEMPLE 16.6 – En reprenant les notations de l'exemple 16.2, on a $\mathrm{Tr}(A) = 0 + 1 = 1$. Pour D, on obtient

$$\mathrm{Tr}(D) = \frac{1+\sqrt{5}}{2} + \frac{1-\sqrt{5}}{2} = 1.$$

On obtient le même résultat, et ce n'est pas un hasard.

LEMME 16.7 – *La trace possède les propriétés suivantes :*

1. *Si M et N sont carrées, alors* $\mathrm{Tr}(MN) = \mathrm{Tr}(NM)$.
2. *Si A et B sont conjuguées, alors* $\mathrm{Tr}(A) = \mathrm{Tr}(B)$.

Démonstration. Pour le (1), on fait un calcul direct. Si $M = (m_{ij})_{i,j}$ et $N = (n_{k\ell})_{k,\ell}$, on trouve en fait

$$\mathrm{Tr}(MN) = \sum_{i,k} m_{ik}n_{ki} = \mathrm{Tr}(NM).$$

Pour le (2), supposons que $B = P^{-1}AP$, et posons $M = P^{-1}$, puis $N = AP$. Alors

$$\mathrm{Tr}(B) = \mathrm{Tr}(MN) = \mathrm{Tr}(NM) = \mathrm{Tr}(APP^{-1}) = \mathrm{Tr}(A).$$

□

EXEMPLE 16.8 – Nous allons pouvoir donner des exemples de matrices qui ne sont *pas* diagonalisables. Commençons par

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Supposons que A soit diagonalisable, donc qu'il existe une matrice inversible P telle que

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = D.$$

On doit alors avoir $\det(A) = 1 = \det(D) = \lambda_1\lambda_2$, et $\mathrm{Tr}(A) = 2 = \mathrm{Tr}(D) = \lambda_1 + \lambda_2$. Nous connaissons donc la somme et le produit de λ_1 et λ_2, et il est alors facile de trouver ces nombres : l'astuce habituelle est de regarder le polynôme

$$(X-\lambda_1)(X-\lambda_2) = X^2-(\lambda_1+\lambda_2)X+\lambda_1\lambda_2 = X^2-2X+1 = (X-1)^2.$$

On en déduit que $\lambda_1 = \lambda_2 = 1$. Mais alors, la matrice D n'est autre que la matrice identité ! Par suite

$$A = PDP^{-1} = P\,\mathrm{Id}\,P^{-1} = PP^{-1} = \mathrm{Id},$$

ce qui est absurde puisque $A \neq \mathrm{Id}$. Cette contradiction montre que P ne peut pas exister, c'est-à-dire que A n'est pas diagonalisable.

Voyons maintenant la matrice

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Celle-ci est-elle diagonalisable ? De nouveau, supposons qu'il existe P telle que $P^{-1}AP = D$, avec D ayant les coefficients λ_1 et λ_2 sur la diagonale, comme ci-dessus. Cette fois-ci, on doit avoir $\lambda_1\lambda_2 = \det(A) = 1$ et $\lambda_1 + \lambda_2 = \mathrm{Tr}(A) = 0$. Ceci donne

$$(X-\lambda_1)(X-\lambda_2) = X^2 + 1.$$

Les nombres λ_1 et λ_2, qui sont des éléments de $\mathbb{K}$, doivent donc être les racines du polynôme $X^2 + 1$. Si $\mathbb{K} = \mathbb{R}$, nous avons déjà une contradiction, puisque les racines sont i et $-i$, qui ne sont pas dans $\mathbb{R}$: on dit que A n'est pas diagonalisable « sur $\mathbb{R}$ ».

Mais on peut considérer A comme une matrice de $M_2(\mathbb{C})$, à coefficients complexes, et rechercher P également à coefficients complexes. Dans ce cas on peut prendre

$$P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \text{ et alors } P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Donc A est diagonalisable sur $\mathbb{C}$. Là encore, nous allons expliquer dans la suite du chapitre comment trouver cette matrice P, que nous avons sortie de nulle part.

voir les exercices 3401, 3404, 5627

La proposition que voici donne des exemples de matrices conjuguées, et en un sens elle les donne même tous.

Proposition 16.9 – *Soit* E *un espace vectoriel de dimension finie, et soit* $f : \mathrm{E} \to \mathrm{E}$ *une application linéaire. Soient* $\mathscr{A}$ *et* $\mathscr{B}$ *deux bases de* E. *Les matrices de* f *dans ces bases respectives sont notées*

$$\mathrm{A} = {}_{\mathscr{A}}[f]^{\mathscr{A}} \quad \text{et} \quad \mathrm{B} = {}_{\mathscr{B}}[f]^{\mathscr{B}}\,.$$

Alors A *et* B *sont conjuguées, et plus précisément on a même* $\mathrm{B} = \mathrm{P}^{-1}\mathrm{AP}$ *où* P *est la matrice de passage*

$$\mathrm{P} = {}_{\mathscr{A}}\mathrm{P}^{\mathscr{B}}\,.$$

Réciproquement, si C *est une matrice de la même taille que* A, *telle que* A *et* C *sont conjuguées, alors il existe une base* $\mathscr{C}$ *de* E *telle que*

$$\mathrm{C} = {}_{\mathscr{C}}[f]^{\mathscr{C}}\,.$$

Démonstration. C'est la formule du changement de base (proposition 15.30), qui affirme précisément que

$${}_{\mathscr{B}}[f]^{\mathscr{B}} = {}_{\mathscr{B}}\mathrm{P}^{\mathscr{A}}{}_{\mathscr{A}}[f]^{\mathscr{A}}{}_{\mathscr{A}}\mathrm{P}^{\mathscr{B}}\,.$$

Rappelons que si $\mathrm{P} = {}_{\mathscr{A}}\mathrm{P}^{\mathscr{B}}$, alors $\mathrm{P}^{-1} = {}_{\mathscr{B}}\mathrm{P}^{\mathscr{A}}$.

Pour la réciproque, soit P telle que $\mathrm{C} = \mathrm{P}^{-1}\mathrm{AP}$. Soient $e_1, e_2, \ldots, e_n$ les vecteurs de E dont les coordonnées dans la base $\mathscr{A}$ sont les colonnes de la matrice P. Puisque P est inversible, ses colonnes forment une base de $\mathbb{K}^n$ (corollaire 14.21), donc la famille $\mathscr{C} = e_1, \ldots, e_n$ est une base de E (proposition 14.29). D'après la formule du changement de base on a

$${}_{\mathscr{C}}[f]^{\mathscr{C}} = {}_{\mathscr{C}}\mathrm{P}^{\mathscr{A}}{}_{\mathscr{A}}[f]^{\mathscr{A}}{}_{\mathscr{A}}\mathrm{P}^{\mathscr{C}} = \mathrm{P}^{-1}\mathrm{AP} = \mathrm{C}. \qquad \square$$

Conclusion : deux matrices sont conjuguées exactement lorsqu'elles représentent la même application linéaire dans deux bases différentes. De nouveau, les techniques matricielles et le concept d'application linéaire vont s'enrichir mutuellement.

Pour diagonaliser, il va être très utile de réfléchir en termes d'applications linéaires. En effet :

Proposition 16.10 – *Soit* $\mathrm{A} \in \mathrm{M}_n(\mathbb{K})$, *et soit* $f : \mathbb{K}^n \to \mathbb{K}^n$ *l'application linéaire définie par* A, *c'est-à-dire* $f(v) = \mathrm{A}v$. *Alors* A *est diagonalisable* $\Longleftrightarrow$ *il existe une base* $e_1, e_2, \ldots, e_n$ *de* $\mathbb{K}^n$ *avec la propriété que* $f(e_i) = \lambda_i e_i$ *pour un certain scalaire* $\lambda_i \in \mathbb{K}$.

Lorsque c'est le cas, soit P *la matrice dont les colonnes sont les vecteurs* e_i ; *on a alors*

$$\mathrm{P}^{-1}\mathrm{AP} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \qquad (*)$$

Démonstration. Soit $\mathscr{C}$ la base canonique de $\mathbb{K}^n$, de sorte que $\mathrm{A} = {}_{\mathscr{C}}[f]^{\mathscr{C}}$. Supposons que la base $\mathscr{B} = e_1, \ldots, e_n$ existe avec la propriété ci-dessus, alors par définition même de la matrice d'une application linéaire, on a

$${}_{\mathscr{B}}[f]^{\mathscr{B}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \qquad (**)$$

Mais d'après la formule du changement de base, on a ${}_{\mathscr{B}}[f]^{\mathscr{B}} = \mathrm{P}^{-1}\mathrm{AP}$ avec $\mathrm{P} = {}_{\mathscr{C}}[f]^{\mathscr{B}}$. Donc (*) est vérifiée, et A est diagonalisable.

Réciproquement, si (*) est vérifiée, on procède à l'envers : on appelle $e_1, \ldots, e_n$ les colonnes de P, qui forment une base $\mathscr{B}$ puisque P est inversible ; la formule du changement de base nous dit que (**) est vérifiée ; et par définition même cela signifie que $f(e_i) = \lambda_i e_i$. $\square$

De nouveau, ces choses portent des noms classiques :

Définition 16.11 – Soit $f : \mathrm{E} \to \mathrm{E}$ une application linéaire. Un *vecteur propre* de f est un vecteur $v \neq 0$ tel que $f(v) = \lambda v$ pour un certain scalaire $\lambda \in \mathbb{K}$. On dit que v et λ sont associés.

Lorsque $\lambda \in \mathbb{K}$ est associé à au moins un vecteur propre, on dit que c'est une *valeur propre* de f.

Enfin, la valeur propre λ étant fixée, l'ensemble des $v \in \mathrm{E}$ tels que $f(v) = \lambda v$ est appelé *l'espace propre* associé à λ. Nous le noterons E_λ.

Notez bien la condition $v \neq 0$, qui est essentielle ; elle garantit que v détermine λ, puisque $\lambda v = \mu v$ entraîne bien $\lambda = \mu$, lorsque $v \neq 0$.

Exemple 16.12 – Prenons deux vecteurs $e_1, e_2 \in \mathbb{R}^2$ qui forment une base, et posons $\mathrm{U} = \mathcal{V}ect(e_1)$, puis $\mathrm{V} = \mathcal{V}ect(e_2)$, de sorte que $\mathbb{R}^2 = \mathrm{U} \oplus \mathrm{V}$. Soit maintenant s la symétrie par rapport à U, dans la direction V.

Par définition, on a $s(e_1) = e_1$, donc e_1 est un vecteur propre de s associé à la valeur propre 1. De même $s(e_2) = -e_2$, donc e_2 est un vecteur propre de s associé à la valeur propre -1. Enfin, en écrivant $\mathscr{B} = e_1, e_2$, on a

$${}_{\mathscr{B}}[s]^{\mathscr{B}} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

et cette matrice est diagonale. Par suite, la matrice de s dans n'importe quelle base est de la forme

$$\mathrm{P}^{-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\mathrm{P},$$

c'est-à-dire qu'elle est diagonalisable.

Nous pouvons complètement terminer les calculs présentés dans l'introduction de ce chapitre :

Exemple 16.13 – Retournons à la matrice

$$\mathrm{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

de l'introduction (et de l'exemple 16.2). Pour trouver (seuls !) une matrice P telle que

$$\mathrm{P}^{-1}\mathrm{AP} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

on utilise d'abord la même astuce que dans l'exemple 16.8 : on doit avoir $\lambda_1 + \lambda_2 = \mathrm{Tr}(\mathrm{A}) = 1$ et $\lambda_1\lambda_2 = \det(\mathrm{A}) = -1$. Donc

$$(\mathrm{X} - \lambda_1)(\mathrm{X} - \lambda_2) = \mathrm{X}^2 - (\lambda_1 + \lambda_2)\mathrm{X} + \lambda_1\lambda_2 = \mathrm{X}^2 - \mathrm{X} - 1\,.$$

Ainsi, λ_1 et λ_2 sont les racines du polynôme $\mathrm{X}^2 - \mathrm{X} - 1$; supposons qu'on les ait numérotées de la façon suivante :

$$\lambda_1 = \frac{1 + \sqrt{5}}{2}, \qquad \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

On comprend déjà mieux d'où provenaient ces $\sqrt{5}$!

Maintenant, soit f l'application $f(v) = \mathrm{A}v$, comme dans la proposition 16.10. Cherchons les vecteurs propres associés à λ_1 : c'est un calcul de système linéaire. En effet, l'équation $f(v) = \lambda_1 v$ s'écrit

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda_1\begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{en posant} \quad v = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Ceci donne le système

$$\begin{cases} -\lambda_1 x & + & y & = & 0 \\ x & + & (1 - \lambda_1)y & = & 0 \end{cases},$$

et en faisant $\mathrm{L}_2 \leftarrow \mathrm{L}_2 + \frac{1}{\lambda_1}\mathrm{L}_1$ on obtient $0 = 0$ sur la deuxième ligne. L'ensemble des solutions est donc décrit par la seule équation $-\lambda_1 x + y = 0$, on peut prendre y comme paramètre, et on constate que l'on a un espace vectoriel de dimension 1, avec comme base par exemple (en prenant $y = 1$) le vecteur

$$e_1 = \begin{pmatrix} \frac{1}{\lambda_1} \\ 1 \end{pmatrix} = \begin{pmatrix} -\lambda_2 \\ 1 \end{pmatrix}.$$

En procédant de la même manière pour λ_2, on montre que l'espace propre associé, c'est-à-dire l'ensemble des vecteurs tels que $f(v) = \lambda_2 v$, est un espace de dimension 1 avec pour base le vecteur

$$e_2 = \begin{pmatrix} -\lambda_1 \\ 1 \end{pmatrix}.$$

Il se trouve que e_1, e_2 est une base de $\mathbb{R}^2$, comme on le voit tout de suite. Si alors P est la matrice dont les colonnes sont e_1 et e_2, la proposition 16.10 nous dit que $\mathrm{P}^{-1}\mathrm{AP}$ est la matrice diagonale annoncée. Voici comment on était venu à bout de la suite de Fibonacci.

voir les exercices 1605, 2467

Nous savons désormais diagonaliser les matrices 2×2, du moins lorsque c'est possible, en procédant comme nous l'avons fait dans l'exemple 16.13. Ce qui semble nous empêcher de faire de même avec des matrices quelconques, c'est que l'on ne sait pas *a priori* quelles sont les valeurs propres potentielles, alors que pour les 2×2 on exploite l'astuce de calcul $(X-\lambda_1)(X-\lambda_2) = X^2 - \mathrm{Tr}(A)X + \det(A)$. Il se trouve qu'il existe un polynôme jouant le même rôle pour les matrices de n'importe quelle taille.

Proposition 16.14 – *Soit* $A \in M_n(\mathbb{K})$, *et soit* f *l'application* $f(v) = Av$. *Alors* λ *est valeur propre de* $f \iff \det(A-\lambda\mathrm{Id}) = 0$.

L'expression $\det(A-\lambda\mathrm{Id})$ est un polynôme en λ, que l'on appelle le *polynôme caractéristique de* A (ou de f). On le note χ_A, ou χ_f.

Démonstration. Cette démonstration est facile en soi, mais il est intéressant de noter la quantité de résultats non-triviaux des chapitres précédents qui entrent en jeu.

Pour $\lambda \in \mathbb{K}$, notons $f-\lambda\mathrm{Id}$ l'application définie par $(f-\lambda\mathrm{Id})(v) = f(v)-\lambda v$. Alors par définition λ est une valeur propre de $f \iff$ il existe $v \neq 0$ tel que $(f-\lambda\mathrm{Id})(v) = 0 \iff \ker(f-\lambda\mathrm{Id}) \neq \{0\}$.

D'après la proposition 15.39, cette condition équivaut à dire que $f-\lambda\mathrm{Id}$ n'est pas injective. D'après le théorème du rang (ou plus précisément le corollaire 15.44), ceci équivaut encore à dire que $f-\lambda\mathrm{Id}$ n'est pas bijective.

D'après la proposition 15.34, ceci revient à affirmer que la matrice de $f-\lambda\mathrm{Id}$ dans la base canonique n'est pas inversible. Or cette matrice est $A-\lambda\mathrm{Id}$, et finalement la proposition 13.5 affirme que cette condition se ramène à $\det(A-\lambda\mathrm{Id}) = 0$. □

Exemple 16.15 – Prenons une matrice 2×2 :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Le polynôme caractéristique est alors

$$\chi_A = \det(A-\lambda\mathrm{Id}) = \det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc),$$

ce qui donne dans ce cas particulier

$$\chi_A = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

On retrouve donc le polynôme de degré 2 qui était apparu dans nos calculs avec les matrices 2×2.

Exemple 16.16 – Prenons maintenant

$$A = \begin{pmatrix} 6 & 2 & 1 \\ -6 & -1 & -2 \\ 0 & 0 & 3 \end{pmatrix}.$$

Le polynôme caractéristique est donné par :

$$\chi_A = \begin{vmatrix} 6-\lambda & 2 & 1 \\ -6 & -1-\lambda & -2 \\ 0 & 0 & 3-\lambda \end{vmatrix} = (3-\lambda)\begin{vmatrix} 6-\lambda & 2 \\ -6 & -1-\lambda \end{vmatrix} = (3-\lambda)(\lambda^2-5\lambda+6) = -(\lambda-3)^2(\lambda-2).$$

Les valeurs propres sont donc 2 et 3, et on dit que 3 a une « multiplicité » de 2 puisque le polynôme caractéristique a $(\lambda-3)^2$ en facteur.

Examinons les vecteurs propres. Pour trouver ceux associés à la valeur propre 2, on résout comme d'habitude le système $Av = 2v$. Faites le calcul, vous trouverez un espace de dimension 1, avec pour base par exemple

$$e_1 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}.$$

Pour la valeur propre 3, on résout $Av = 3v$. L'ensemble des solutions est de dimension 2, avec pour base par exemple

$$e_2 = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \quad \text{et} \quad e_3 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}.$$

(Vérifiez ceci.)

Il se trouve que e_1, e_2, e_3 est une base de $\mathbb{R}^3$. Nous avons donc une base de vecteurs propres, ce qui signifie que A est diagonalisable. Plus précisément, si P est la matrice dont les colonnes sont e_1, e_2, e_3, on sait sans calcul supplémentaire que

$$P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Donnons quelques proriétés générales du polynômes caractéristique.

Proposition 16.17 – *Soient* A *et* B *des matrices de* $M_n(\mathbb{K})$.

1. *Si* A *et* B *sont conjuguées, alors* $\chi_A = \chi_B$.
2. *Si* A *est diagonalisable, alors son polynôme caractéristique* χ_A *est scindé sur* $\mathbb{K}$.

Rappelons qu'un polynôme en λ est dit *scindé* si c'est un produit de facteurs de degré 1, c'est-à-dire s'il est de la forme

$$c(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n).$$

Démonstration. Si $B = P^{-1}AP$ alors

$$P^{-1}(A-\lambda\mathrm{Id})P = P^{-1}AP - \lambda P^{-1}P = B - \lambda\mathrm{Id},$$

donc $B-\lambda\mathrm{Id}$ et $A-\lambda\mathrm{Id}$ sont conjuguées. Elles ont donc le même déterminant, ce qui donne le (1).

Pour le (2), on utilise le (1) en prenant B diagonale. On a alors

$$\chi_A = \chi_B = \begin{vmatrix} \lambda_1-\lambda & 0 & \cdots & 0 \\ 0 & \lambda_2-\lambda & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n-\lambda \end{vmatrix} = (\lambda_1-\lambda)\cdots(\lambda_n-\lambda).$$

voir les exercices 2762, 2763

Donc χ_A est bien scindé dans ce cas. □

Nous avons essentiellement décrit toutes les étapes nécessaires pour diagonaliser une matrice. Mais on peut faire une remarque supplémentaire qui va nous simplifier la tâche.

Après avoir trouvé des vecteurs propres pour les diverses valeurs propres, nous devons vérifier si l'on peut trouver une base complète, formée de ces vecteurs propres. Il s'ensuit un travail de vérification, pour savoir si certaines familles sont libres. Vous aurez peut-être remarqué que dans les exemples jusqu'à présent, on n'avait jamais de mauvaise surprise : ce n'est pas un hasard, comme nous allons le montrer.

Commençons par

Lemme 16.18 – *Soit $f\colon \mathrm{E} \to \mathrm{E}$ une application linéaire, et soit e_1, …, e_n une famille de vecteurs propres de f. On suppose que e_i est associé à λ_i, et que les nombres λ_1, …, λ_n sont distincts.*

Alors la famille $e_1, \ldots, e_n$ est libre.

Démonstration. Par récurrence sur n, le résultat étant évident pour $n = 1$ (un vecteur propre est non-nul par définition).

Supposons donc que l'on ait $n+1$ vecteurs propres, et une combinaison linéaire nulle, disons

$$\alpha_1 e_1 + \cdots + \alpha_{n+1} e_{n+1} = 0. \qquad (*)$$

Appliquons f aux deux membres de (*) ; en utilisant $f(e_i) = \lambda_i e_i$, il vient

$$\alpha_1 \lambda_1 e_1 + \cdots + \alpha_{n+1} \lambda_{n+1} e_{n+1} = 0. \qquad (**)$$

Multiplions (*) par λ_{n+1}, et retranchons le résultat à (**) ; il vient

$$(\lambda_1 - \lambda_{n+1})\alpha_1 e_1 + \cdots + (\lambda_n - \lambda_{n+1})\alpha_n e_n = 0.$$

Par récurrence, tous les coefficients de cette combinaison linéaire sont nuls, donc $(\lambda_i - \lambda_{n+1})\alpha_i = 0$ pour $1 \leq i \leq n$. Comme $\lambda_i - \lambda_{n+1} \neq 0$ par hypothèse, on en tire $\alpha_i = 0$ pour ces valeurs de i. Ensuite il est clair que $\alpha_{n+1} = 0$ également, et la famille est donc libre. □

Du coup, l'entreprise de diagonalisation s'en trouve simplifiée : en deux mots, lorsque l'on réunit des bases des différents espaces propres, on obtient une famille qui est automatiquement libre. Si elle comporte suffisamment de vecteurs, et seulement dans ce cas, on a réussi à diagonaliser. Plus précisément :

Proposition 16.19 – *Soit $\mathrm{A} \in \mathrm{M}_n(\mathbb{K})$, et soit f l'application linéaire associée. Soient $\lambda_1, \lambda_2, \ldots, \lambda_m$ les racines distinctes du polynôme caractéristique de A.*

Pour chaque λ_i, soit n_i la dimension de l'espace propre associé $\ker(f - \lambda_i \,\mathrm{Id})$, et soit

$$e_{i1}, e_{i2}, \ldots, e_{in_i}$$

une base de cet espace.

Alors A est diagonalisable si et seulement si

$$\sum_{i=1}^{m} n_i = n.$$

Lorsque c'est le cas, la famille comprenant tous les vecteurs e_{ij} est une base de $\mathbb{K}^n$, formée de vecteurs propres de f.

Démonstration. Montrons que la famille formée de tous les e_{ij} est libre. Si on a une combinaison linéaire nulle de la forme

$$\sum_{i,j} \alpha_{i,j} e_{ij} = 0,$$

alors on pose $e_i = \sum_j \alpha_{i,j} e_{ij}$. On a $f(e_i) = \lambda_i e_i$ (puisque chaque e_{ij} est un vecteur propre associé à λ_i). De plus on a $e_1 + e_2 + \cdots + e_m = 0$.

Cette relation donnerait une contradiction au lemme précédent, à moins que tous les vecteurs e_i soient nuls (et ne sont donc pas des vecteurs propres). On a donc $e_i = 0$ et donc chaque $\alpha_{i,j} = 0$ puisque la famille $e_{i1}, e_{i2}, \ldots, e_{in_i}$ est libre.

Comme annoncé, la famille formée de tous les e_{ij} est libre. Elle comporte $\sum_i n_i$ élements, donc si $\sum_i n_i = n = \dim \mathbb{K}^n$, c'est une base. Dans ce cas, on est en présence d'une base formée de vecteurs propres, donc A est diagonalisable.

Pour finir, supposons que $\sum_i n_i < n$. Un vecteur propre v de f doit appartenir à un $\ker(f - \lambda_i \,\mathrm{Id})$ pour un certain i, donc en particulier $v \in \mathcal{V}ect(e_{ij})$. Mais l'espace $\mathcal{V}ect(e_{ij})$ est de dimension $\sum_i n_i < n$, et il ne peut pas contenir une base de $\mathbb{K}^n$. Donc il n'existe pas de base de $\mathbb{K}^n$ formée de vecteurs propres de f, et A n'est pas diagonalisable. □

En particulier, on a le résultat suivant, étonnamment simple :

Corollaire 16.20 – *Soit $\mathrm{A} \in \mathrm{M}_n(\mathbb{K})$. Si le polynôme caractéristique de A possède n racines distinctes dans $\mathbb{K}$, alors A est diagonalisable.*

Démonstration. Soient $\lambda_1, \ldots, \lambda_n$ les valeurs propres (distinctes). Chaque espace $\ker(f - \lambda_i \,\mathrm{Id})$ est $\neq \{0\}$, par définition, donc il est de dimension ≥ 1. Ainsi, la somme $\sum_i n_i$ de la proposition précédente est $\geq n$; mais bien sûr cette somme est aussi $\leq n$ puisque l'on a vu que c'est le nombre de vecteurs dans une certaine famille libre.

Finalement $\sum_i n_i = n$, donc A est diagonalisable (et de plus chaque $n_i = 1$). □

Lorsque l'on souhaite montrer qu'une matrice est diagonalisable, mais que l'on n'a pas besoin de trouver expressément les vecteurs propres, ce corollaire est évidemment idéal. Nous verrons une application dans le chapitre sur les équations différentielles.

Avant de donner des exemples d'utilisation de ces derniers résultats, nous allons résumer la méthode développée dans ce chapitre.

Soit $A \in M_n(\mathbb{K})$, soit $f : \mathbb{K}^n \to \mathbb{K}^n$ l'application linéaire définie par $f(v) = Av$. Pour tenter de diagonaliser A, on suit les étapes suivantes.

1. On calcule le polynôme caractéristique $\chi_A = \det(A - \lambda \mathrm{Id})$, et on trouve ses racines distinctes $\lambda_1, \ldots, \lambda_m$ dans $\mathbb{K}$.
   - Si χ_A n'est pas scindé, alors A n'est pas diagonalisable, et on s'arrête.
   - Si χ_A est scindé, on passe à l'étape suivante. Si $n = m$, c'est-à-dire si on a n valeurs propres distinctes, alors on sait déjà que A est diagonalisable.
2. Pour chaque λ_i, on calcule une base de $\ker(f - \lambda_i \mathrm{Id})$. On en déduit sa dimension n_i.
   - Si $\sum_i n_i < n$, la matrice A n'est pas diagonalisable, et on s'arrête.
   - Si $\sum_i n_i = n$, la matrice est diagonalisable, et on passe à l'étape suivante.
3. Soit $e_{i1}, e_{i2}, \ldots, e_{in_i}$ une base de $\ker(f - \lambda_i \mathrm{Id})$. Alors la réunion de tous ces vecteurs est une base de $\mathbb{K}^n$. Soit P la matrice dont les colonnes sont, dans l'ordre :

$$e_{11}, \ldots, e_{1n_1}, \ldots, e_{m1}, \ldots, e_{mn_m}.$$

Alors sans calcul supplémentaire on sait que

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 & \cdots \\ \vdots & \ddots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_2 & 0 & 0 & \cdots \\ \vdots & 0 & 0 & 0 & \ddots & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \lambda_2 & \cdots \\ \vdots & 0 & 0 & 0 & \vdots & \vdots & \ddots \end{pmatrix}.$$

À droite on a une matrice diagonale, où λ_1 apparaît n_1 fois, puis λ_2 apparaît n_2 fois, etc.

Exemple 16.21 – Prenons

$$A = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 3 & 2 \\ 3 & 1 & 1 \end{pmatrix}.$$

Le calcul du polynôme caractéristique donne

$$\chi_A = -\lambda^3 + 5\lambda^2 + 2\lambda - 24.$$

Selon la façon de calculer le déterminant, ce polynôme peut vous apparaître factorisé, ce qui est évidemment une bonne chose pour calculer les racines. En règle générale, les opérations sur les lignes ou les colonnes, plutôt que les développements, ont tendance à produire des polynômes factorisés.

Mais admettons que nous ayons obtenu le polynôme sous la forme ci-dessus. Il faut trouver une racine « évidente ». Dans cette situation, il est souvent utile de revenir à la matrice : n'est-il pas clair que si on ajoute 2 sur la diagonale de A, alors la première colonne devient égale à la troisième ? Donc $\det(A + 2\mathrm{Id}) = 0$, ce qui signifie que -2 est valeur propre, et $\chi_A(-2) = 0$. On termine ensuite facilement la factorisation :

$$\chi_A = -(\lambda + 2)(\lambda^2 - 7\lambda + 12) = -(\lambda + 2)(\lambda - 4)(\lambda - 3).$$

Les valeurs propres sont -2, 4 et 3. On a trois valeurs propres distinctes, donc on sait que la matrice est diagonalisable. (Ceci conclut l'étape 1).

Avant de poursuivre les calculs, dressons la liste de ce que nous pouvons déjà prévoir. Nous allons trouver des vecteurs e_1, e_2 et e_3, vecteurs propres associés à -2, 4 et 3 respectivement ; ces vecteurs vont former une base de $\mathbb{R}^3$, automatiquement. Soit P la matrice dont les colonnes sont e_1, e_2, e_3. Alors

$$P^{-1}AP = \begin{pmatrix} -2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Ce sont les conclusions de l'étape 3. Si les valeurs propres n'avaient pas été distinctes, on n'aurait pas pu prévoir le résultat de l'étape 2.

Cette étape 2 reste à faire, de toute façon. Pour la valeur propre -2 par exemple, on cherche $\ker(f + 2\mathrm{Id})$ ce qui revient à résoudre

$$\begin{cases} 3x - y + 3z = 0 \\ 2x + 5y + 2z = 0 \\ 3x + y + 3z = 0 \end{cases}$$

En quelques étapes on constate que ce système équivaut aux équations $y = 0$ et $x + z = 0$. En prenant $z = 1$ par exemple, on obtient

$$e_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

De la même manière, on obtient

$$e_2 = \begin{pmatrix} -1 \\ 5 \\ 1 \end{pmatrix}, \qquad e_3 = \begin{pmatrix} 1 \\ 12 \\ 5 \end{pmatrix}.$$

voir les exercices 2468, 2580, 2764, 2765, 2766

– Fin de la première lecture –

Nous l'avons vu, la diagonalisation ne fonctionne pas toujours. À défaut de pouvoir diagonaliser, on tente parfois de *trigonaliser*.

Définition 16.22 – Une matrice carrée est dite *triangulaire* (supérieure) lorsque les coefficients sous la diagonale sonts nuls.

On dit qu'une matrice carrée A est *trigonalisable* lorsqu'elle est conjuguée à une matrice triangulaire, c'est-à-dire lorsqu'il existe P telle que $P^{-1}AP$ est triangulaire.

Exemple 16.23 – La matrice

$$A = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$$

est triangulaire (donc trigonalisable !). On peut voir facilement qu'elle n'est pas diagonalisable : en effet $\chi_A = (\lambda - 3)^2$, donc la seule valeur propre est 3, et l'espace propre correspondant n'est que de dimension 1.

Il est quand même possible de faire des calculs avec A, par exemple de calculer A^n, même si c'est plus difficile que pour une matrice diagonale. Posons

$$D = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \quad \text{et} \quad N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

de sorte que $A = D + N$. On a $N^2 = 0$, ce qui va grandement nous aider. Notons également que $DM = MD = 3M$ pour toute matrice M. On peut donc calculer

$$A^2 = (D+N)^2 = (D+N)(D+N) = D^2+ND+DN+N^2 = D^2+6N.$$

De même

$$A^3 = (D^2 + 6N)(D + N) = D^3 + 6ND + D^2N + 6N^2$$
$$= D^3 + 18N + 9N = D^3 + 27N.$$

Visiblement A^n est de la forme $D^n + a_nN$. On a $a_2 = 6 = 2\times 3$ et $a_3 = 27 = 3 \times 3^2$. Supposons que $a_n = n3^{n-1}$, alors

$$A^{n+1} = (D^n + n3^{n-1}N)(D + N) = D^{n+1} + (n+1)3^nN.$$

Par récurrence, ceci montre que $a_n = n3^{n-1}$ pour tout n. Finalement la matrice A^n vaut

$$D^n + n3^{n-1}N = \begin{pmatrix} 3^n & 0 \\ 0 & 3^n \end{pmatrix} + n3^{n-1}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3^n & n3^{n-1} \\ 0 & 3^n \end{pmatrix}.$$

Lorsqu'une matrice est triangulaire, avec $\lambda_1, \lambda_2, \ldots, \lambda_n$ sur la diagonale, son polynôme caractéristique est $(\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$. En particulier il est scindé ; toute matrice dont le polynôme caractéristique n'est pas scindé ne peut donc pas être trigonalisable.

Sur $\mathbb{C}$, tous les polynômes sont scindés d'après le théorème fondamental de l'algèbre, donc on ne risque pas de trouver de matrices non-trigonalisables en procédant comme ça. Et en fait, il n'y en a pas ! En effet :

Proposition 16.24 – *Soit* $A \in M_n(\mathbb{C})$. *Alors A est trigonalisable.*

Démonstration. On va procéder par récurrence sur n (pour $n = 1$ il n'y a rien à prouver).

Soit χ_A le polynôme caractéristique de A. Comme on est sur $\mathbb{C}$, ce polynôme possède au moins une racine, disons λ_1. Si f est l'application $f(v) = Av$, alors par définition, on a donc un vecteur propre e_1 tel que que $f(e_1) = \lambda_1 e_1$.

D'après le théorème de la base incomplète, on peut trouver une base $\mathscr{B} = e_1, e_2, \ldots, e_n$ de $\mathbb{C}^n$ dont le premier vecteur est e_1. La matrice de f dans la base $\mathscr{B}$ est de la forme

$$_{\mathscr{B}}[f]^{\mathscr{B}} = \begin{pmatrix} \lambda_1 & * \; * \; \cdots \; * \\ 0 & \\ \vdots & A' \\ 0 & \end{pmatrix},$$

où A′ est une matrice $(n-1)\times(n-1)$ (et $*$ est un coefficient quelconque).

Par récurrence, il existe une matrice inversible Q telle que la matrice $T' = Q^{-1}A'Q$ est triangulaire. Posons alors

$$P = \begin{pmatrix} 1 & 0 \; \cdots \; 0 \\ 0 & \\ \vdots & Q \\ 0 & \end{pmatrix}.$$

L'inverse de P est de la même forme, avec Q remplacée par Q^{-1}. Un petit calcul montre alors que

$$P^{-1}\,{}_{\mathscr{B}}[f]^{\mathscr{B}}\,P = \begin{pmatrix} \lambda_1 & * \; * \; \cdots \; * \\ 0 & \\ \vdots & Q^{-1}A'Q = T' \\ 0 & \end{pmatrix}.$$

En particulier cette matrice est triangulaire, puisque T′ l'est ; notons-la T, de sorte que $_{\mathscr{B}}[f]^{\mathscr{B}} = PTP^{-1}$.

Enfin, notons $\mathscr{C}$ la base canonique, ce qui permet d'écrire $A = {}_{\mathscr{C}}[f]^{\mathscr{C}}$. Notons également $R = {}_{\mathscr{B}}P^{\mathscr{C}}$. La formule du changement de base nous dit que

$$A = R^{-1}\,{}_{\mathscr{B}}[f]^{\mathscr{B}}\,R = R^{-1}PTP^{-1}R = S^{-1}TS,$$

avec $S = P^{-1}R$. Ainsi, la matrice A est bien conjuguée à la matrice triangulaire T. □

Le théorème montre donc qu'il existe P telle que

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \cdots & * \\ & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix}.$$

Le polynôme caractéristique de A est donc

$$(\lambda_1 - \lambda)\cdots(\lambda_n - \lambda),$$

et les nombres λ_i sont les valeurs propres de A. On observe alors que $\lambda_1 + \lambda_2 + \cdots + \lambda_n = \mathrm{Tr}(A)$, et $\lambda_1\lambda_2\cdots\lambda_n = \det(A)$, formules que nous connaissions pour $n = 2$. En d'autres termes :

Corollaire 16.25 – *La somme des valeurs propres d'une matrice complexe, comptées avec leurs multiplicités, est égale à sa trace ; le produit de ces valeurs propres est égal à son déterminant.*

On a une notion évidente de matrice *triangulaire inférieure*, lorsque les coefficients au-dessus de la diagonale sont nuls. On déduit du théorème précédent que :

Corollaire 16.26 – *Toute matrice complexe est conjuguée à une matrice triangulaire inférieure.*

Démonstration. On applique le théorème à la matrice transposée ${}^tA$: il existe donc P telle que $T = P^{-1}({}^tA)P$ est triangulaire supérieure. En transposant, on obtient ${}^tP\,A\,{}^tP^{-1} = {}^tT$. Posant $Q = {}^tP^{-1}$, on a bien montré que $Q^{-1}AQ = {}^tT$ était triangulaire inférieure. □

À l'aide de la proposition 16.24, on peut montrer que « la plupart » des matrices sont diagonalisables. Plus précisément, nous allons montrer :

PROPOSITION 16.27 – *Soit* A *une matrice à coefficients complexes. Alors il existe une suite* $(A_n)_{n\geq 0}$ *de matrices diagonalisables, telle que*

$$A_n \xrightarrow[n\to\infty]{} A.$$

Rappelons un peu ce que la convergence signifie. Si $A \in M_N(\mathbb{C})$, alors chaque $A_n \in M_N(\mathbb{C})$; dire que A_n converge vers A signifie que les N^2 coefficients de A_n convergent les N^2 coefficients de A. Mais on peut identifier une matrice de $M_N(\mathbb{C})$ avec un vecteur de $\mathbb{C}^{N^2}$ (ou encore de $\mathbb{R}^{4N^2}$, en voyant $\mathbb{C}$ comme $\mathbb{R}^2$), et penser à la convergence en termes de norme comme dans la proposition 3.29, si l'on préfère.

Cette proposition est à rapprocher de ce que font les ordinateurs lorsqu'ils effectuent des calculs numériques approchés. Étant donnée une matrice à coefficients réels ou complexes, et quelle que soit la précision requise, un ordinateur (auquel on ne demande pas explicitement de faire des calculs exacts) va toujours annoncer qu'elle est diagonalisable (sur $\mathbb{C}$). Concrètement, les valeurs propres, dont seule une valeur approchée sera calculée, seront toujours considérées comme distinctes, et dans ce cas on a le corollaire 16.20.

C'est cette même idée qui va guider la démonstration.

Démonstration. D'après la proposition 16.24, on peut trouver P telle que $P^{-1}AP$ est triangulaire ; utilisons les notations suivantes :

$$T = P^{-1}AP = \begin{pmatrix} \lambda_1 & * & * & * \\ & \lambda_2 & \cdots & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix}. \qquad (\dagger)$$

Soit maintenant T_n obtenue en prenant les coefficients de T mais avec les changements suivants sur la diagonale :

$$T_n = \begin{pmatrix} \lambda_1 + \frac{1}{n} & * & * & * \\ & \lambda_2 + \frac{2}{n} & \cdots & \\ & & \ddots & \\ & & & \lambda_N + \frac{N}{n} \end{pmatrix}.$$

(Les coefficients $*$ sont les mêmes que dans $(\dagger)$; seule la diagonale est changée.) Donc sur la ligne i on rencontre le coefficient $\mu_i = \mu_{i,n} = \lambda_i + \frac{i}{n}$.

Nous allons voir que les coefficients μ_i sont tous distincts, pour n suffisamment grand. En effet, si i et j sont deux indices distincts, alors nous avons deux cas à considérer : d'abord, si $\lambda_i = \lambda_j$, alors $\mu_i - \mu_j = \frac{i-j}{n} \neq 0$ puisque $i \neq j$; si par contre $\lambda_i \neq \lambda_j$, alors puisque μ_i converge vers λ_i, et μ_j converge vers λ_j, il est clair que $\mu_i \neq \mu_j$ dès que n est suffisamment grand.

D'après le corollaire 16.20, la matrice T_n est diagonalisable, pour tous les n suffisamment grands. Or il est clair que

$$T_n \xrightarrow[n\to\infty]{} T \quad \text{et que} \quad PT_nP^{-1} \xrightarrow[n\to\infty]{} PTP^{-1} = A.$$

Enfin, nous notons que PT_nP^{-1} est diagonalisable pour n suffisamment grand, comme T_n. □

Quatrième partie

Appendices

Annexe A

Les mathématiques au lycée

Dans cet appendice nous allons recenser les résultats que vous avez probablement vus au lycée et qui vont être utiles cette année. Le programme officiel de terminale (et d'avant) change souvent, et vous pouvez donc vous attendre à découvrir une chose ou deux, mais rien de très surprenant.

Certains élèves vont souhaiter commencer par une lecture détaillée de cet appendice avant même de lire le premier chapitre, alors que d'autres préfèreront remettre cette annexe à plus tard. L'auteur pense que le plus simple est de consulter l'appendice à mesure que le texte principal fait référence à des choses du lycée. (Pour prendre un exemple simple : mieux vaut lire la définition d'une « bijection » dans le chapitre 1 avant de se replonger dans la définition de arccos.) Mais tout est possible.

Vous serez peut-être déstabilisés par le fait que dans cet appendice, nous parlons de dérivées, de racines n-ièmes, de l'exponentielle, etc, alors qu'en lisant les chapitres du livre en commençant par le début, on a l'impression de reprendre tout à zéro. Pourquoi ça ? Pourquoi ne pas *continuer* le cours là où il s'est arrêté en terminale ?

La raison en est, ne le prenez pas mal, que les connaissances que vous avez acquises en mathématiques jusqu'ici ont quelques imperfections. C'est une chose naturelle et souhaitable : une première exposition aux concepts des mathématiques se doit de repousser quelques difficultés à plus tard. Pensez aux dérivées : savez-vous pourquoi une fonction dont la dérivée est positive est nécessairement croissante ? Pouvez-vous écrire une démonstration *complète* ? Et pourtant, les dérivées sont déjà de l'histoire un peu ancienne pour vous. (Il y a bien pire : savez-vous définir ce qu'est un angle ? Savez-vous montrer qu'il y a un nombre dont le carré vaut 2 ?)

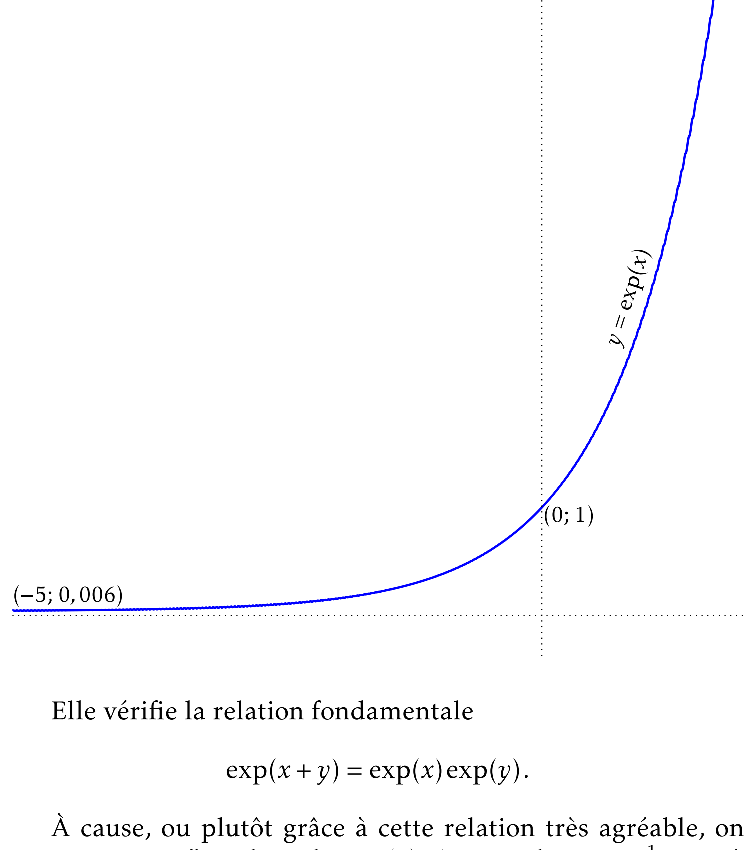
Cette année, tout en abordant un grand nombre de nouveaux thèmes, nous allons combler toutes ces petites lacunes. Le plus simple pour organiser l'exposition est finalement de repartir du début, en allant vite sur les concepts qui vous sont déjà familiers.

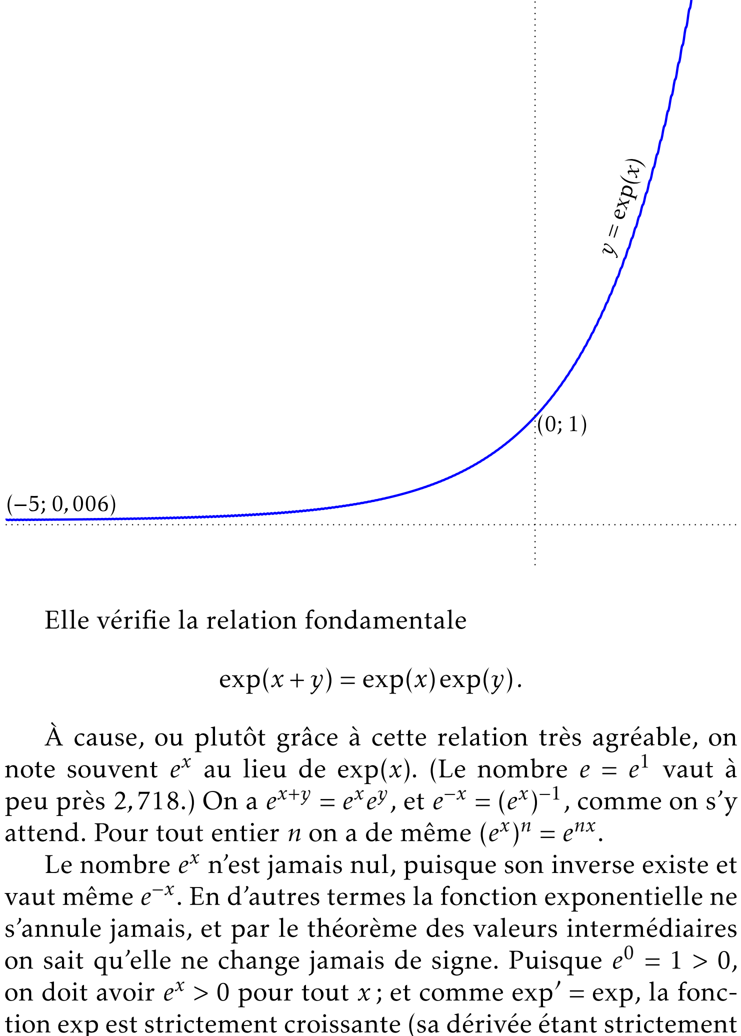
Pour les raisons que nous venons de donner, cet appendice est rédigé dans un style assez différent de celui des chapitres, qui contiennent le cours à proprement parler. Pour le moment, on suppose que les fonctions, les dérivées, les angles, $\sqrt{2}\ldots$, sont des choses que vous connaissez, et nous dressons la liste des choses que vous savez en dire.

Symbole	Nom	Symbole	Nom
α	alpha	ξ	xi (*ksi*)
β	bêta	o	omicron
γ	gamma	π	pi
δ	delta	ρ	rhô
ε	epsilon	σ	sigma
ζ	zêta	τ	tau
θ	thêta	υ	upsilon
ι	iota	ϕ	phi
κ	kappa	χ	khi
λ	lambda	ψ	psi
μ	mu	ω	omega
ν	nu		

Le terme de « fonctions usuelles » désigne un arsenal de fonctions que l'homme honnête se doit de connaître, tant elles reviennent souvent dans les problèmes que l'on rencontre.

Exponentielle & Logarithme. Un résultat fondamental affirme qu'il existe une fonction $\exp\colon \mathbb{R} \to \mathbb{R}$, appelée *l'exponentielle*, qui vérifie $\exp(0) = 1$ et qui est égale à sa propre dérivée : $\exp' = \exp$. (Et en fait il n'y en a qu'une : voir l'encadré « Quelques démonstrations » pour plus d'informations.) L'allure de l'exponentielle est donnée sur le dessin suivant.





Elle vérifie la relation fondamentale

$$\exp(x+y) = \exp(x)\exp(y).$$

À cause, ou plutôt grâce à cette relation très agréable, on note souvent e^x au lieu de $\exp(x)$. (Le nombre $e = e^1$ vaut à peu près 2,718.) On a $e^{x+y} = e^x e^y$, et $e^{-x} = (e^x)^{-1}$, comme on s'y attend. Pour tout entier n on a de même $(e^x)^n = e^{nx}$.

Le nombre e^x n'est jamais nul, puisque son inverse existe et vaut même e^{-x}. En d'autres termes la fonction exponentielle ne s'annule jamais, et par le théorème des valeurs intermédiaires on sait qu'elle ne change jamais de signe. Puisque $e^0 = 1 > 0$, on doit avoir $e^x > 0$ pour tout x ; et comme $\exp' = \exp$, la fonction exp est strictement croissante (sa dérivée étant strictement positive), conformément au dessin.

L'exponentielle réalise une bijection $\exp\colon \mathbb{R} \to \mathbb{R}^{>0} =]0;+\infty[$. La bijection réciproque s'appelle le *logarithme népérien* et est notée $\ln\colon \mathbb{R}^{>0} \to \mathbb{R}$. On a la relation fondamentale $\exp(\ln(x)) = x$ pour tout $x > 0$, et $\ln(\exp(x)) = x$ pour tout x. De plus

$$\ln(ab) = \ln(a) + \ln(b),$$

puisque les deux membres de cette équation deviennent tous les deux ab en prenant l'exponentielle. Notez aussi $\ln(1/a) = -\ln(a)$, ainsi que les valeurs particulières $\ln(1) = 0$, $\ln(e) = 1$. L'allure du logarithme est donnée sur le dessin ci-dessous (avec l'exponentielle en pointillés).

Conformément à ce que l'on voit sur ces images, on a

$$\lim_{x\to-\infty} e^x = 0, \qquad \lim_{x\to+\infty} e^x = +\infty,$$

ainsi que

$$\lim_{x\to 0} \ln(x) = -\infty, \qquad \lim_{x\to+\infty} \ln(x) = +\infty.$$

D'ailleurs « l'exponentielle l'emporte sur les polynômes », ce qui signifie que les limites ci-dessus « l'emportent » dans les formules que voici :

$$\lim_{x\to-\infty} x^n e^x = 0, \qquad \lim_{x\to+\infty} \frac{e^x}{x^n} = +\infty,$$

où n est un entier ≥ 0.

Cosinus & Sinus. Soit C du cecle de centre $(0,0)$ et de rayon 1, dans le plan. Considérons l'intersection de ce cercle et de la demi-droite qui fait un angle θ avec l'axe horizontal ; nous obtenons un point dont les coordonnées sont $(\cos(\theta), \sin(\theta))$, par définition (c'est-à-dire que les nombres $\cos(\theta)$ et $\sin(\theta)$ sont définis par cette situation géométrique). On prononce « cosinus θ » et « sinus θ ».

Listons quelques unes des propriétés essentielles de ces fonctions. Avant tout, puisqu'un angle de 2π représente un tour complet, on note que la demi-droite faisant un angle θ avec l'horizontale est confondue avec la demi-droite faisant un angle $\theta + 2\pi$ avec l'horizontale : donc $\cos(\theta + 2\pi) = \cos(\theta)$ et $\sin(\theta + 2\pi) = \sin(\theta)$. On dit que les fonctions cosinus et sinus sont 2π-*périodiques*.

Ensuite, remarquons qu'un point (x, y) appartient à C s'il est à distance 1 de $(0,0)$, donc si $x^2 + y^2 = 1$. En particulier on a toujours

$$\cos(\theta)^2 + \sin(\theta)^2 = 1.$$

Les dérivées des fonctions cosinus et sinus sont données par

$$\cos' = -\sin, \qquad \sin' = \cos.$$

Par ailleurs on a les relations à savoir

$$\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y),$$

$$\sin(x+y) = \cos(x)\sin(y) + \sin(x)\cos(y).$$

On peut retrouver toutes les autres formules « de trigonométrie » à partir de ces deux-là, par exemple $\sin(x + \frac{\pi}{2}) = \cos(x)\sin(\frac{\pi}{2}) + \sin(x)\cos(\frac{\pi}{2}) = \cos(x)$. On peut essayer de se rappeler qu'en faisant $x = y$ on a aussi

$$\cos(2x) = 2\cos(x)^2 - \sin(x)^2 = 2\cos(x)^2 - 1,$$

et

$$\sin(2x) = 2\sin(x)\cos(x).$$

Et enfin, signalons les relations

$$\cos(-x) = \cos(x),, \qquad \sin(-x) = -\sin(x),$$

qui sont évidentes géométriquement (mais vous pouvez essayer de les démontrer, à titre d'exercice, de manière algébrique).

Le dessin suivant donne l'allure de nos deux fonctions (cosinus en trait plein, sinus en pointillés).

Pour regarder de plus près le cosinus par exemple, on peut se contenter de faire une étude sur $[-\pi, \pi]$, par périodicité ; et comme cos est paire, on peut même se restreindre à $[0, \pi]$. Sur cet intervalle, le cosinus réalise une bijection $\cos\colon [0,\pi] \to [-1,1]$, qui est décroissante. Sa bijection réciproque, appelée « arccosinus » et notée $\arccos\colon [-1,1] \to [0,\pi]$, est représentée sur le dessin suivant.

On a donc les relations $\cos(\arccos(x)) = x$ pour tout $x \in [-1,1]$, et $\arccos(\cos(\theta)) = \theta$ pour $\theta \in [0,\pi]$ (attention, cette relation n'est pas valide pour $\theta \in \mathbb{R}$ quelconque !).

Avec le sinus, on montre que c'est sur $[-\frac{\pi}{2}, \frac{\pi}{2}]$ que l'on a une bijecion $\sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1,1]$. La bijection réciproque s'appelle « arcsinus », elle est notée $\arcsin\colon [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$. La situation est résumée sur le dessin suivant.

On a $\sin(\arcsin(x)) = x$ pour tout $x \in [-1,1]$ et $\arcsin(\sin(\theta)) = \theta$ pour tout $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ (mais pas pour tout $\theta \in \mathbb{R}$).

Les dérivées sont données par

$$\arccos'(x) = -\frac{1}{\sqrt{1-x^2}}, \qquad \arcsin'(x) = \frac{1}{\sqrt{1-x^2}}.$$

Notez que l'on a alors $(\arccos(x) + \arcsin(x))' = 0$, donc la quantité $(\arccos(x) + \arcsin(x))$ est constante ; en prenant $x = 0$ on découvre que

$$\arccos(x) = \frac{\pi}{2} - \arcsin(x),$$

pour tout $x \in [-1,1]$.

Quelques démonstrations

Il est très utile de savoir qu'il n'y a *qu'une seule* fonction avec les propriétés de l'exponentielle. Nous pouvons même le démontrer facilement, en se basant uniquement sur les règles de calcul des dérivées.

Soit donc f telle que $f' = f$ et $f(0) = 1$. On regarde $g(x) = \exp(-x)f(x)$ et l'on trouve $g'(x) = -\exp(-x)f(x) + \exp(-x)f'(x) = 0$. Donc g est constante, en tant que fonction dont la dérivée est nulle, disons $g(x) = c$. Puisque $f(0) = 1$, alors $g(0) = 1 = c$, et on en tire $\exp(-x)f(x) = 1$; ceci montre que l'on n'a aucun choix pour $f(x)$, il faut prendre $f(x) = 1/\exp(-x)$, et on a bien l'unicité. Au passage, en prenant $f(x) = \exp(x)$ on (re)découvre que $\exp(-x) = \exp(x)^{-1}$ pour tout nombre x.

Appliquons ceci à la fonction $f(x) = \exp(-y)\exp(x+y)$, où y est un nombre fixé. On a bien $f' = f$, et $f(0) = \exp(-y)\exp(y) = 1$, donc par unicité $f(x) = \exp(x)$. En multipliant par $\exp(y)$ nous découvrons la relation fondamentale :

$$\exp(x+y) = \exp(x)\exp(y).$$

Pour calculer la dérivée du logarithme, on dérive les deux membres de l'égalité $\exp(\ln(x)) = x$, on obtient avec la règle habituelle $\exp(\ln(x)) \cdot \ln'(x)) = 1 = x\ln'(x)$, et donc

$$\ln'(x) = \frac{1}{x}.$$

Passons au cosinus et au sinus. Notons donc que l'on a

$$\cos'' = -\cos, \qquad \sin'' = -\sin.$$

Il y a ici un résultat d'unicité qui ressemble un peu à ce que l'on a dit sur l'exponentielle, mais plus compliqué : si f est une fonction telle que $f'' = -f$, alors on a

$$f(x) = A\cos(x) + B\sin(x)$$

où A et B sont des constantes. On ne peut pas démontrer ça avec les méthodes du lycée, mais vous utilisez probablement ce résultat en Physique (en rapport avec les oscillateurs). Nous le verrons dans le chapitre « Équations différentielles ».

Si l'on considère par exemple $f(x) = \cos(x + \frac{\pi}{2})$, alors on a bien $f''(x) = -f(x)$, donc f est de la forme ci-dessus. En prenant $x = 0$, on obtient

$$f(0) = \cos(\frac{\pi}{2}) = 0 = A \times 1 + B \times 0 = A,$$

donc $A = 0$, et en $x = \frac{\pi}{2}$ on a de même

$$f(\frac{\pi}{2}) = \cos(\pi) = -1 = B,.$$

et l'on découvre la formule $\cos(x + \frac{\pi}{2}) = -\sin(x)$.

Recommençons avec $f(x) = \cos(x+y)$, où y est un nombre fixé. Avec la même méthode on obtient $A = \cos(y)$ et $B = \cos(y + \frac{\pi}{2})$, donc $B = -\sin(y)$. Ceci donne

$$\cos(x+y) = \cos(x)\cos(y) - \sin(x)\sin(y),$$

formule fondamentale. De la même manière on montre que

$$\sin(x+y) = \cos(x)\sin(y) + \sin(x)\cos(y).$$

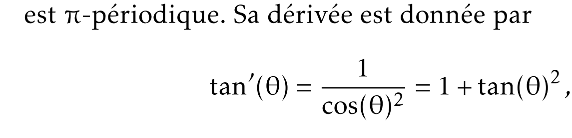
Tangente. On pose

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$$

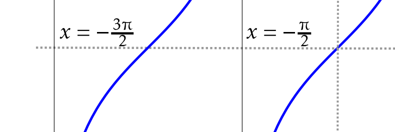
lorsque ça a un sens, c'est-à-dire pour $\theta \in \mathbb{R} \smallsetminus \{\frac{\pi}{2} + k\pi,\ k \in \mathbb{Z}\}$. Puisque $\sin(\theta + \pi) = -\sin(\theta)$ et $\cos(\theta + \pi) = -\cos(\theta)$, on a $\tan(\theta + \pi) = \tan(\theta)$: en d'autres termes, la fonction tangente est π-périodique. Sa dérivée est donnée par

$$\tan'(\theta) = \frac{1}{\cos(\theta)^2} = 1 + \tan(\theta)^2,$$

ce qui montre que la fonction est croissante *sur chaque intervalle* où elle est définie. L'allure de la tangente est donnée sur le dessin suivant.



La restriction à $]-\frac{\pi}{2}, \frac{\pi}{2}[$ de la fonction tangente est une bijection sur $\mathbb{R}$ tout entier. La bijection réciproque $\arctan\colon \mathbb{R} \to]-\frac{\pi}{2}, \frac{\pi}{2}[$ s'appelle « arctangente », et son graphe est donné ci-dessous. On a donc $\tan(\arctan(x)) = x$ pour tout $x \in \mathbb{R}$, et $\arctan(\tan(\theta)) = \theta$ pour $\theta \in]-\frac{\pi}{2}, \frac{\pi}{2}[$ (mais pas pour tout $\theta \in \mathbb{R}$).



Nous avons rappelé les dérivées des fonctions usuelles. Rappelons également que la dérivée de $x \mapsto x^n$ est $x \mapsto nx^{n-1}$, pour tout entier $n \geq 0$. Partant de là, on se sert constamment des règles suivantes :

- ◇ $(f+g)' = f' + g'$.
- ◇ $(fg)' = f'g + fg'$.
- ◇ $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$.

EXEMPLE A.1 – Calculons par exemple

$$\left(\frac{\cos(x)}{3+5x^2}\right)' = \frac{-\sin(x)(3+5x^2) - 10x\cos(x)}{(3+5x^2)^2}.$$

Il y a enfin une règle pour calculer la dérivée de $x \mapsto f(g(x))$. Curieusement, elle semble disparaître peu à peu des programmes des lycées. La voici :

- ◇ $f(g(x))' = f'(g(x))g'(x)$.

EXEMPLE A.2 – La dérivée de $\sin(4x)$ est donc $4\cos(4x)$, comme on le voit avec cette formule pour $g(x) = 4x$ et $f(x) = \sin(x)$. De même la dérivée de $\exp(2x+1)$ est $2\exp(2x+1)$. En général la dérivée de $\exp(g(x))$ est $g'(x)\exp(g(x))$. Par contre

$$\ln(g(x))' = \frac{g'(x)}{g(x)}$$

et donc la dérivée de $\ln(12x)$ est $\frac{1}{x}$ (peu surprenant dans la mesure où $\ln(12x) = \ln(12) + \ln(x)$).

EXEMPLE A.3 – La dérivée de $\exp(\alpha\ln(x))$ est donc $\frac{\alpha}{x}\exp(\alpha\ln(x))$. Ce résultat mérite quelques commentaires.

La quantité $\exp(\alpha\ln(x))$ est souvent notée x^α (pour $x > 0$ et α quelconque), ce qui est cohérent avec la notation pour les entiers : c'est-à-dire que si α est un entier ≥ 0, alors la définition de x^α que l'on vient de donner revient à $x \times x \times \cdots \times x$, le tout α fois. De même si α est un entier ≤ 0, alors $x^\alpha = (\frac{1}{x})^{-\alpha}$ comme prévu.

On a la relation $x^\alpha x^\beta = x^{\alpha+\beta}$ comme on le vérifie facilement, et aussi $(x^\alpha)^\beta = x^{\alpha\beta}$. Par exemple $(x^{\frac{1}{n}})^n = x$, et $x^{1/n}$ est bien une « racine n-ième » de x.

La formule pour la dérivée stipule alors que $(x^\alpha)' = \alpha x^{\alpha-1}$. Autrement dit : la formule $(x^n)' = nx^{n-1}$ marche aussi pour les puissances qui ne sont pas des nombres entiers. Déjà pour $\alpha = -1$ on retrouve que la dérivée de $\frac{1}{x} = x^{-1}$ est $-x^{-2} = \frac{-1}{x^2}$ (on peut aussi retrouver ça par la formule générale pour la dérivée de f/g, évidemment). Dans le même genre, la dérivée de $x^{\frac{1}{n}}$ est $\frac{1}{n}x^{\frac{1-n}{n}}$.

EXEMPLE A.4 – Il faut savoir dériver la composée de trois fonctions, c'est-à-dire une expression de la forme $f(g(h(x)))$. D'après la formule précédente, la dérivée est

$$f'(g(h(x)))g(h(x))' = f'(g(h(x)))g'(h(x))h'(x).$$

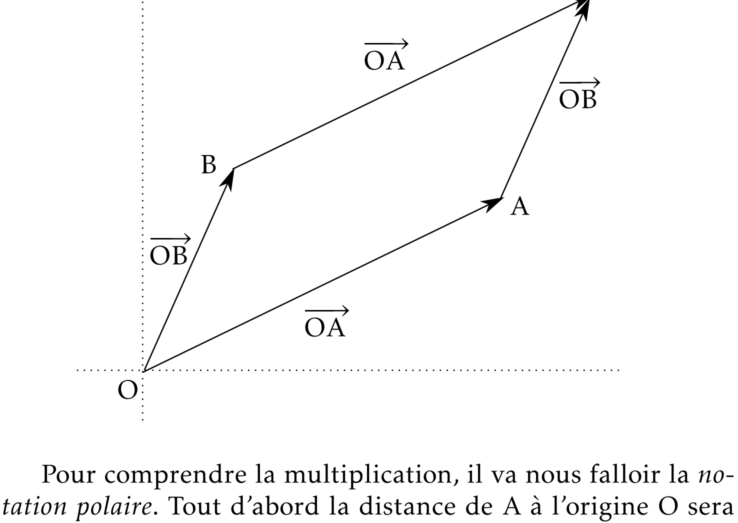
Par exemple

$$\exp(\cos(x^2+1))' = -2x\exp(\cos(x^2+1))\sin(x^2+1).$$

Au lycée on vous a parlé de l'existence des *nombres complexes*, qui sont de la forme $x+iy$, où $x, y \in \mathbb{R}$ et i est un nombre complexe vérifiant $i^2 = -1$. Leur ensemble est noté $\mathbb{C}$. Dans le cours nous reviendrons sur leurs propriétés algébriques (en particulier nous montrerons que ces nombres existent bien, ce qui n'est pas complètement évident pour i !). Pour l'instant nous souhaitons faire quelques rappels sur l'utilisation des complexes en géométrie.

On va toujours identifier $\mathbb{C}$ et $\mathbb{R}^2$, le nombre $x+iy$ étant identifié à (x, y). Un nombre complexe $A = x+iy$ peut donc être vu comme un point dans le plan, d'abscisse x et d'ordonnée y. On dira parfois que $x+iy$ est *l'affixe* du point (x, y).

Les opérations algébriques sur les complexes peuvent être interprétées géométriquement. C'est peut-être la différence $B-A$ qui se voit le mieux : l'abscisse et l'ordonnée de $B-A$ sont les coordonnées du *vecteur* $\overrightarrow{AB}$, qui donne la translation menant de A à B. D'ailleurs en écrivant O pour le point d'affixe nulle (l'origine), on peut penser au nombre complexe A comme représentant le vecteur $\overrightarrow{OA}$. L'addition se comprend alors simplement, puisque $A+B$ est obtenu en opérant une translation de vecteur $\overrightarrow{OA}+\overrightarrow{OB}$, appliquée à l'origine. En d'autres termes, $A+B$ est comme dans le dessin suivant.



Pour comprendre la multiplication, il va nous falloir la *notation polaire*. Tout d'abord la distance de A à l'origine O sera notée $|A|$, que l'on appelle le *module de* A. Si $A = x+iy$, on a ainsi $|A| = \sqrt{x^2+y^2}$. Il en résulte $|A \cdot B| = |A| \cdot |B|$.

On va maintenant écrire $e^{i\theta}$ pour

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Cette écriture est l'une des plus mystérieuses de la terminale. Plus loin dans ce livre nous donnerons des explications complètes (voir le chapitre « L'exponentielle »). Pour l'instant on va se contenter de remarquer que les formules pour $\cos(a+b)$ et $\sin(a+b)$ montrent que

$$e^{i(a+b)} = e^{ia}e^{ib}.$$

C'est une excellente nouvelle : il est bien plus facile de mémoriser $e^{i(a+b)} = e^{ia}e^{ib}$ que de mémoriser séparément les deux formules pour cosinus et sinus, alors qu'on peut les retrouver en examinant les parties réelles et imaginaires. Notons aussi que le module de $e^{i\theta}$ est 1, puisque $\cos(\theta)^2 + \sin(\theta)^2 = 1$.

Supposons maintenant que le point $A = x+iy$ est sur le cercle de centre O et de rayon 1 ; en d'autres termes, $x^2+y^2 = 1$. Comme nous l'avons dit dans notre étude des fonctions cosinus et sinus ci-dessus, on peut écrire $x = \cos(\theta)$ et $y = \sin(\theta)$ où θ est l'angle entre l'axe horizontal et $\overrightarrow{OA}$. On a donc, avec notre nouvelle notation, $A = e^{i\theta}$. On dit parfois que θ est *l'argument* de A ; puisque θ n'est pas unique (il est défini à un multiple de 2π près), on devrait dire *un* argument de A.

Si maintenant A est un complexe quelconque (mais non nul), le nombre $A/|A|$ est de module 1, donc on peut écrire $A/|A| = e^{i\theta}$ comme ci-dessus, et θ est encore l'angle entre l'axe horizontal et $\overrightarrow{OA}$. Finalement

$$A = |A|\,e^{i\theta}.$$

Nous pouvons enfin examiner le produit de $A = |A|e^{i\theta}$ par $B = |B|e^{i\theta'}$: nous obtenons $AB = |A||B|e^{i(\theta+\theta')}$. Plus visuellement : les modules se multiplient, les arguments s'ajoutent.

Pour énoncer ceci de manière encore plus géométrique, fixons $A = e^{i\theta} \in \mathbb{C}$, un complexe de module 1, et considérons l'application du plan vers lui-même donnée par $Z \mapsto AZ$. D'après le calcul précédent, il s'agit de *la rotation de centre* O *et d'angle* θ. De même, si $\rho > 0$ est réel, l'application $Z \mapsto \rho Z$ est *l'homothétie de centre* O *et de rapport* ρ. On peut donc affirmer en général que, pour $A = \rho e^{i\theta}$ un complexe non-nul quelconque, l'application $Z \mapsto AZ$ est la composée d'une rotation et d'une homothétie, comme indiqué.

Un raisonnement similaire montre que la rotation de centre quelconque M et d'angle θ s'exprime avec les complexes par la formule $Z \mapsto M + e^{i\theta}(Z-M)$. (On pourrait presque écrire $Z \mapsto M + e^{i\theta}\overrightarrow{MZ}$, encore que multiplier un vecteur par un complexe peut prêter à confusion, et n'est pas recommandé.) L'homothétie de centre M et de rapport ρ est donnée par $Z \mapsto M + \rho(Z-M)$. La composée des deux transformations que nous venons d'indiquer est $Z \mapsto M + \rho e^{i\theta}(Z-M)$.

Soient maintenant A, B et C trois points du plan. Dire que l'angle entre $\overrightarrow{AB}$ et $\overrightarrow{AC}$ est θ revient précisément à dire que l'image de B par la rotation de centre A et d'angle θ, suivie de l'homothétie de centre A et de rapport $\rho = |C-A|/|B-A|$, est tout simplement C. En clair

$$C = A + \rho e^{i\theta}(B-A).$$

Plus algébriquement : on écrit le complexe $\frac{C-A}{B-A}$ sous la forme « polaire » $\rho e^{i\theta}$, et alors θ est l'angle (défini à un multiple de 2π près, comme d'habitude) entre $\overrightarrow{AB}$ et $\overrightarrow{AC}$. En coordonnées, écrivons $B-A = x+iy$ (ou $\overrightarrow{AB} = (x, y)$, ce qui revient au même), et $C-A = x'+iy'$. Alors

$$\rho e^{i\theta} = \frac{C-A}{B-A} = \frac{x'+iy'}{x+iy} = \frac{xx'+yy'}{x^2+y^2} + i\,\frac{xy'-x'y}{x^2+y^2}$$

(en multipliant numérateur et dénominateur par $x-iy$). En comparant les parties réelles, sachant que $\rho = |C-A|/|B-A|$, on obtient

$$xx'+yy' = |C-A||B-A|\cos(\theta).$$

La quantité $xx'+yy'$ est appelée le *produit scalaire* des deux vecteurs $\overrightarrow{AB}$ et $\overrightarrow{AC}$. Il existe plusieurs notations, comme $\overrightarrow{AB} \cdot \overrightarrow{AC}$, ou encore $\left(\overrightarrow{AB}, \overrightarrow{AC}\right)$. La dernière formule affirme que le produit scalaire est le produit des normes des deux vecteurs, que multiplie le cosinus de l'angle entre eux.

Deux vecteurs sont perpendiculaires, par définition, lorsque l'angle entre eux est $\pm\frac{\pi}{2}$ (à un multiple de 2π près). Puisque $e^{i\frac{\pi}{2}} = i$, les formules ci-dessus montre que $\overrightarrow{AB}$ et $\overrightarrow{AC}$ sont perpendiculaires exactement lorsque

$$\frac{C-A}{B-A} = \rho i,$$

c'est-à-dire lorsque le rapport $\frac{C-A}{B-A}$ est imaginaire pur. Et surtout, en termes du produit scalaire, les deux vecteurs sont perpendiculaires lorsque $\overrightarrow{AB} \cdot \overrightarrow{AC} = 0$. Cette définition très algébrique de la notion de perpendicularité est utilisée sans cesse.

Jouons un peu avec l'écriture « polaire ». On a l'identité $(e^{i\theta})^n = e^{ni\theta}$. On en déduit que

$$\left(e^{\frac{2ik\pi}{n}}\right)^n = e^{2ik\pi} = 1\,,$$

donc que les nombres $e^{\frac{2ik\pi}{n}}$ sont des « racines n-ièmes » du nombre 1 : on dit souvent « racines n-ièmes de l'unité ». Ceci nous donne n nombres différents, pour $0 \leq k \leq n-1$ (ensuite pour $k = n$ on obtient le même nombre que pour $k = 0$, etc). On peut montrer qu'il n'y en a pas d'autres, et nous venons de donner la description des n racines n-ièmes de l'unité. (Une démonstration de ceci, ainsi que bien d'autres informations, sera donnée avec le corollaire B.15).

Si maintenant z est un complexe quelconque non-nul, écrivons $z = \rho e^{i\theta}$. Nous connaissons le nombre $\rho^{\frac{1}{n}} = e^{\frac{1}{n}\ln(\rho)}$, qui vérifie $\left(\rho^{\frac{1}{n}}\right)^n = \rho$. Si on pose alors $w = \rho^{\frac{1}{n}} e^{\frac{i\theta}{n}}$, on a donc $w^n = z$: tout nombre complexe possède une racine n-ième.

Par ailleurs, les n nombres $w_k = e^{\frac{2ik\pi}{n}} w$ pour $0 \leq k \leq n-1$ vérifient tous $w_k^n = z$, et on montre qu'il n'y en a pas d'autres. En d'autres termes, tout nombre complexe non-nul possède exactement n racines n-ièmes, données par la recette ci-dessus.

Exemple A.5 – Prenons un exemple archi-simple : quelles sont les racines 4-èmes de $1+i$? Nous allons avoir besoin des racines 4-ièmes de l'unité, qui sont

$$e^{\frac{2ik\pi}{4}} = 1, i, -1, -i \quad \text{pour} \quad k = 0, 1, 2, 3\,.$$

Ensuite on calcule le module de $z = 1 + i$, qui vaut $|z| = \sqrt{2}$. Le nombre $\frac{z}{|z|} = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$ doit pouvoir se mettre sous la forme $e^{i\theta}$. Il n'y a pas vraiment de meilleure méthode que de reconnaître que $\theta = \frac{\pi}{4}$ convient. Finalement $z = \sqrt{2}e^{\frac{i\pi}{4}}$, et une racine 4-ième possible est (en gardant en tête que $\sqrt{2} = 2^{\frac{1}{2}}$) le nombre $w = 2^{\frac{1}{8}} e^{\frac{i\pi}{16}}$.

Au total, les 4 racines 4-ièmes sont $w, iw, -w$ et $-iw$.

Sous peu nous allons avoir besoin de la notation concise pour les sommes, à l'aide du symbole Σ. Rappelons le fonctionnement.

Supposons donnés des nombres $a_0, a_1, a_2, \ldots$ Lorsque $n \geq m$ on écrit

$$\sum_{k=m}^{n} a_k \quad \text{pour} \quad a_m + a_{m+1} + a_{m+2} + \cdots + a_n \,.$$

Il est essentiel de comprendre que la lettre k n'a aucune importance et aurait pu être remplacée par n'importe quelle autre : c'est-à-dire que

$$\sum_{k=m}^{n} a_k = \sum_{i=m}^{n} a_i = \sum_{j=m}^{n} a_j = \cdots$$

On dit que i, j et k sont des *variables muettes*.

On utilise souvent des « changements de variables », sous la forme du petit raisonnement suivant (c'est juste un exemple !) : si k varie de m à n, posons $i = k+1$, alors i varie de $m+1$ à $n+1$, et $a_k = a_{i-1}$, d'où

$$\sum_{k=m}^{n} a_k = \sum_{i=m+1}^{n+1} a_{i-1} \,.$$

Puisque les variables sont muettes, on peut donc écrire

$$\sum_{k=m}^{n} a_k = \sum_{k=m+1}^{n+1} a_{k-1} \,.$$

Si cette dernière formule vous paraît moins naturelle que la précédente (avec des i), alors ne l'employez pas, mais il faut être prêt à comprendre ces égalités lorsque quelqu'un d'autre les utilise.

Soit X un ensemble comprenant n éléments. Combien y-a-t-il de sous-ensembles de X ayant k éléments ? Notons ce nombre

$$\binom{n}{k}$$

pour l'instant (en anticipant sur le fait que la réponse ne dépend que de n et k et non pas de X, ce qui va être établi tout de suite). Un petit raisonnement va nous donner une formule simple.

Soit en effet m le nombre de façons de sélectionner k éléments (distincts) dans X et de les numéroter : $x_1, x_2, \ldots, x_k$. Nous allons déterminer m de deux manières différentes. D'un côté, on peut choisir quel élément va être pris pour x_1, et nous avons n choix dans X ; puis pour x_2 il nous reste $n-1$ choix, puisque x_1 est déjà pris ; pour x_3 on a $n-2$ possibilités puisqu'on ne peut plus prendre ni x_1 ni x_2, et ainsi de suite. Ainsi

$$m = n(n-1)(n-2)\cdots(n-k+1).$$

(Pour vous convaincre du $n-k+1$, essayez $k=2$: on a $n(n-1)$ choix, et $n-1 = n-2+1$.)

D'autre part, pour choisir les éléments, on peut d'abord choisir un sous-ensemble de X comprenant k éléments, et il y a $\binom{n}{k}$ façons de le faire ; et ensuite il faut numéroter les éléments (choisir lequel on appelle x_1, lequel est x_2, etc). En raisonnant un peu de la même façon que ci-dessus, on voit qu'il y a $k(k-1)(k-2)\cdots\times 2\times 1$ façons de faire cette numérotation. Au total

$$m = \binom{n}{k} k(k-1)(k-2)\cdots\times 2\times 1.$$

On a bien une formule pour $\binom{n}{k}$, en comparant. On va utiliser la notation *factorielle* :

$$n! = n(n-1)(n-2)\cdots\times 3\times 2\times 1.$$

(Prononcé « factorielle n ».) Nous avons donc

$$m = \frac{n!}{(n-k)!} \quad \text{et} \quad m = \binom{n}{k} k!,$$

d'où

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

À partir de la formule explicite, vous pouvez démontrer immédiatement toutes les identités célèbres, comme

$$\binom{n}{k} = \binom{n}{n-k},$$

ou

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}, \qquad (*)$$

Ceci dit, vous êtes encouragés à démontrer ces égalités en raisonnant sur les sous-ensembles de X, c'est un excellent exercice.

À l'aide de (*), on peut calculer très facilement les nombres ci-dessus. On va construire un tableau, qui contient dans la n-ième ligne et dans la k-ième colonne le nombre $\binom{n}{k}$. Il va être commode d'adopter la convention que $\binom{n}{k} = 0$ si $k > n$ ou si $k < 0$, et on vérifie que (*) est vraie pour toutes les valeurs de k ; par ailleurs, on convient de ne pas écrire les zéros dans le tableau (on laisse un blanc). Les premières lignes sont donc :

1			
1	1		
1	2	1	
1	3	3	1

C'est le *triangle de Pascal* qui apparaît peu à peu. Pour écrire la ligne suivante, on utilise (*) comme ceci :

1			
1	1		
1	2	1	
1	3	3	1
		6	

(Chaque nombre est donc la somme de celui immédiatement au-dessus et de celui à gauche au-dessus ; vérifiez que la ligne suivante est $1, 4, 6, 4, 1$). Le calcul est très rapide, et il suffit de se rappeler que la toute première ligne (pour $n = 0$) contient juste un 1 (pour $k = 0$). On gagne cependant un peu de temps à mémoriser quelques lignes.

L'utilisation la plus fameuse des nombres $\binom{n}{k}$ est la *formule du binôme de Newton*, qui affirme

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

(Ici a et b sont des nombres réels ou complexes, ou en fait n'importe quels éléments avec lesquels les règles de calcul usuelles s'appliquent.) Par exemple on a

$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4,$$

d'après le triangle de Pascal. Précisons qu'un terme de la forme $a^k b^{n-k}$ est appelé un *binôme* (car il fait intervenir deux éléments différents), et que la formule de Newton vaut aux nombres $\binom{n}{k}$ le nom de *coefficients binomiaux*.

Donnons une idée de démonstration, pour commencer. On fait quelques développements « sans permuter a et b » :

$$(a+b)^2 = (a+b)(a+b) = a^2 + ab + ba + b^2.$$

Puis

$$\begin{aligned}(a+b)^3 &= (a^2 + ab + ba + b^2)(a+b) \\ &= a^3 + aba + ba^2 + b^2a + a^2b + ab^2 + bab + b^3.\end{aligned}$$

Ensuite on regroupe les termes, puisque aba, a^2b et ba^2 sont égaux, par exemple. On trouve $a^3 + 3a^2b + 3ab^2 + b^3$, puisqu'il y avait trois termes donnant a^2b, un seul donnant a^3, etc.

Cette petite analyse nous fait réaliser qu'un terme de $(a+b)^n$, après développement mais avant de faire permuter a et b et de regrouper, est de la forme $x_1x_2\cdots x_n$ avec chaque $x_i \in \{a, b\}$; et par ailleurs, chaque écriture $x_1x_2\cdots x_n$ se retrouve une fois et une seule.

Quand on veut regrouper les termes, il faut compter combien de fois on trouve $a^k b^{n-k}$, pour chaque k. Facile : pour obtenir $a^k b^{n-k}$ à partir de $x_1x_2\cdots x_n$, il s'agit de choisir quels indices $i_1, \ldots, i_k$ vérifient $x_{i_1} = x_{i_2} = \cdots = x_{i_k} = a$, et pour les autres indices $x_j = b$. Il faut donc choisir un sous-ensemble de k éléments parmi l'ensemble $\{1, \ldots, n\}$ de tous les indices, donc il y a $\binom{n}{k}$ façons de le faire. Finalement, on obtient un terme $\binom{n}{k} a^k b^{n-k}$, pour chaque k, comme on voulait le montrer.

Cette démonstration est presque rigoureuse, mais certains la trouveront un peu vague tout de même. Pour rédiger un argument inattaquable (mais qui explique moins bien la formule), on va faire une *récurrence*, après quelques rappels.

Le principe de récurrence s'applique lorsque l'on est en présence d'une propriété $P(n)$, dépendant d'un nombre entier n, qui vérifie

1. que $P(0)$ est vraie (c'est *l'initialisation*),
2. que pour tout n, si $P(n)$ est vraie alors $P(n+1)$ est également vraie.

Dans ce cas pour n'importe quel entier n la propriété $P(n)$ est vraie.

C'est une technique de démonstration très courante, nous allons l'appliquer pour établir la formule du binôme de Newton, selon laquelle

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

pour tout entier n. Ici la propriété $P(n)$ est « la formule pour $(a+b)^n$ est vraie pour n ». On vérifie sans peine $P(0)$ qui affirme que $(a+b)^0 = 1$, et aussi $P(1)$ qui affirme que $(a+b)^1 = a+b$, et ci-dessus nous avons vérifié $P(2)$ et $P(3)$. L'initialisation est donc largement faite ($P(0)$ seule aurait suffi).

Supposons donc que $P(n)$ est vraie et penchons nous sur la propriété $P(n+1)$. On écrit

$$\begin{aligned}(a+b)^{n+1} &= (a+b)^n(a+b)\\ &= \left(\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\right)(a+b).\end{aligned}$$

Ici on a utilisé $P(n)$ pour développer $(a+b)^n$. Le reste n'est que calcul, l'étape essentielle étant l'utilisation de la formule (*) pour les coefficients binomiaux :

$$\begin{aligned}(a+b)^{n+1} &= \sum_{k=0}^{n} \binom{n}{k} a^{k+1} b^{n-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k}\\ &= \sum_{k=0}^{n+1} \binom{n}{k-1} a^k b^{n+1-k} + \sum_{k=0}^{n} \binom{n}{k} a^k b^{n+1-k}\\ &= \sum_{k=0}^{n+1} \left(\binom{n}{k-1} + \binom{n}{k}\right) a^k b^{n+1-k}\\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} a^k b^{n+1-k}.\end{aligned}$$

(Notez comment, pour obtenir la deuxième égalité, nous avons remplacé k par $k-1$ dans la somme de gauche ; la somme est inchangée si l'on fait maintenant varier k de 1 à $n+1$ au lieu de 0 à n, mais nous avons même laissé k démarrer à 0 car $\binom{n}{-1} = 0$.) On a bien obtenu $P(n+1)$ en supposant $P(n)$, et par le principe de récurrence, on sait maintenant que $P(n)$ est vraie pour tout n.

Pourquoi le principe de récurrence est-il valide, au fait ? On peut répondre que c'est évident ; c'est une excellente réponse. Mais on peut avoir le désir de montrer que le principe de récurrence se déduit d'un principe encore plus évident (sachant qu'on pourrait continuer sans cesse dans cette direction, voir la fin du chapitre « Ensembles »). Voici par exemple une propriété particulièrement claire : en écrivant $\mathbb{N}$ pour l'ensemble des entiers ≥ 0, alors toute partie $A \subset \mathbb{N}$ possède un plus petit élément, si elle est non-vide.

Si cette propriété de $\mathbb{N}$ vous paraît indiscutable, alors il doit en être de même pour le principe de récurrence. En effet, soit A l'ensemble des entiers n tels que $P(n)$ n'est *pas* vérifiée. Montrons que, si $P(0)$ est vraie, et si $P(n)$ entraîne $P(n+1)$, alors A est vide. S'il n'en était pas ainsi, la partie A aurait un plus petit élément n_0 ; on a $n_0 \neq 0$ car $P(n_0)$ n'est pas vraie ; donc $n_0 - 1$ est un entier ≥ 0, et $n_0 - 1 \notin A$ par minimalité de n_0, donc $P(n_0 - 1)$ est vraie, par définition. On constate que $P(n_0 - 1)$ est vraie mais pas $P(n_0)$, une contradiction qui montre que $A = \emptyset$, autrement dit $P(n)$ est vraie pour tout $n \in \mathbb{N}$.

Soit $S_n = 1+2+\cdots+n$. Il est indispensable de savoir que $S_n = \frac{n(n+1)}{2}$. On peut montrer ça par récurrence très facilement, mais il est bien plus joli de remarquer directement que

$$\begin{array}{rcccccc} S_n + S_n & = & (1+ & 2 & +\cdots+ & n) \\ & + & (n+ & n-1 & +\cdots+ & 1), \end{array}$$

et l'on voit bien que $2S_n$ est la somme de n paquets totalisant à chaque fois $n+1$ (qui vaut aussi $2+(n-1) = 3+(n-2)$ etc). On dit parfois que S_n est la « somme d'une série en progression arithmétique ». À titre d'exercice, trouvez une formule pour $(a+b)+(a+2b)+\cdots+(a+nb)$, quels que soient a et b.

Dans la même veine, vous devez savoir que si $T_n = 1+r+r^2+\cdots+r^n$, alors

$$T_n = \frac{1-r^{n+1}}{1-r}$$

lorsque $r \neq 1$. On parle de la « somme d'une série en progression géométrique ». Pour le montrer, écrire simplement

$$T_n(1-r) = (1+r+\cdots+r^n) - (r+r^2+\cdots r^{n+1}) = 1-r^{n+1}\,.$$

On peut utiliser cette formule pour factoriser $a^{n+1} - b^{n+1}$, pour tous nombres a, b avec $a \neq 0$. En effet

$$a^{n+1} - b^{n+1} = a^{n+1}(1-r^{n+1})$$

avec $r = \frac{b}{a}$. On en tire

$$\begin{aligned} a^{n+1} - b^{n+1} &= a^{n+1}(1-r)(1+r+\cdots+r^n) \\ &= (a-b)(a^n + a^{n-1}b + a^{n-2}b^2 + \cdots ab^{n-1} + b^n)\,. \end{aligned}$$

Annexe B

L'exponentielle

C'est dans cet appendice que nous allons définitivement combler toutes les lacunes qui peuvent subsister dans vos connaissances du lycée. Nous parlons notamment de lacunes *logiques*, qui ne sont pas de votre fait : par exemple, au lycée comme dans l'appendice A on donne une définition du cosinus et du sinus basée sur la notion intuitive d'angle. Mais jamais on ne définit ce qu'est un angle ! Vous allez voir que la façon rigoureuse de s'en sortir est de définir les fonctions trigonométriques d'abord et les angles après. (C'est donc *logiquement* différent.)

Voici le programme. On va définir pour tout complexe $z \in \mathbb{C}$ un autre nombre $e^z \in \mathbb{C}$, à partir « de rien » (mais en faisant appel à la notion de convergence absolue tout de même, ce qui explique que ce n'est pas faisable au lycée). Ensuite on montre que pour $x \in \mathbb{R}$, le nombre e^x est bien celui auquel on pense en terminale ; notre première victoire est donc une vraie définition de la fonction exponentielle usuelle.

Ensuite on définit $\cos(\theta)$ et $\sin(\theta)$ comme étant les parties réelles et imaginaires de $e^{i\theta}$ lorsque $\theta \in \mathbb{R}$, de sorte que l'égalité $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ est vraie par définition. On démontre alors (et c'est difficile) que tout point A sur le cercle de centre $(0,0)$ et de rayon 1 est de la forme $e^{i\theta}$ pour un nombre $\theta \in \mathbb{R}$ défini à un multiple de 2π près – mais surprise, vous verrez que nous *définirons* le nombre π au passage (comme étant le plus petit réel tel que $\cos(\frac{\pi}{2}) = 0$, en gros).

Cette étude étant faite, on peut alors dire que θ est « une mesure de l'angle entre $\overrightarrow{OA}$ et l'horizontale », ce qui définit presque le mot *angle*. Nous disons « presque » parce que c'est seulement la « mesure de l'angle » qui est définie : il n'y a pas du tout de définition mathématique de l'angle lui-même, finalement (on peut se forcer à en donner une, mais d'une manière ou d'une autre on se ramène toujours à la mesure).

Le présent appendice contient également une démonstration du « théorème fondamental de l'algèbre », et quelques résultats sur l'exponentielle de matrice.

Définition B.1 – Soit $z \in \mathbb{C}$. On note $\exp(z)$ ou e^z le nombre

$$e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!} = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{z^k}{k!}.$$

Le nombre complexe e^z est appelé *l'exponentielle* de z.

L'existence de la limite a été montrée dans l'exemple 3.26.
La plus importante propriété de l'exponentielle est sans conteste la suivante :

Théorème B.2 – *Soient a et b deux nombres complexes. Alors*

$$e^{a+b} = e^a e^b.$$

En particulier, pour tout nombre complexe z, on peut calculer $e^z e^{-z} = e^0 = 1$. On constate donc que $e^z \neq 0$.

Démonstration. On a $e^{a+b} = \lim u_n$, en posant

$$u_n = \sum_{k=0}^{n} \frac{(a+b)^k}{k!}.$$

En utilisant la formule de Newton pour le binôme, on peut développer et obtenir :

$$\begin{aligned} u_n &= \sum_{k=0}^{n} \sum_{p=0}^{k} \binom{k}{p} \frac{a^p b^{k-p}}{k!} \\ &= \sum_{k=0}^{n} \sum_{p=0}^{k} \frac{a^p}{p!} \frac{b^{k-p}}{(k-p)!}. \end{aligned}$$

Si maintenant on pose $\alpha_{p,q} = \frac{a^p}{p!} \frac{b^q}{q!}$, et si on note

$$\mathrm{T}_n = \{(p,q) \in \mathbb{N} \times \mathbb{N} \mid 0 \leq p+q \leq n\},$$

alors le calcul précédent s'écrit

$$u_n = \sum_{(p,q)\in \mathrm{T}_n} \alpha_{p,q}.$$

De même, on a $e^a e^b = \lim v_n$ où

$$v_n = \left(\sum_{p=0}^{n} \frac{a^p}{p!} \right) \left(\sum_{q=0}^{n} \frac{b^q}{q!} \right).$$

On peut développer le produit :

$$v_n = \sum_{p=0}^{n} \sum_{q=0}^{n} \frac{a^p}{p!} \frac{b^q}{q!}.$$

Si on pose cette fois-ci

$$\mathrm{C}_n = \{(p,q) \in \mathbb{N} \times \mathbb{N} \mid 0 \leq p \leq n \text{ et } 0 \leq q \leq n\},$$

alors on a

$$v_n = \sum_{(p,q)\in \mathrm{C}_n} \alpha_{p,q}.$$

On souhaite montrer que (u_n) et (v_n) ont la même limite, ou encore que $v_n - u_n \xrightarrow[n\to\infty]{} 0$. Cette différence s'écrit

$$v_n - u_n = \sum_{(p,q)\in \mathrm{C}_n \smallsetminus \mathrm{T}_n} \alpha_{p,q}.$$

On a donc la majoration

$$|v_n - u_n| \leq \sum_{(p,q)\in \mathrm{C}_n \smallsetminus \mathrm{T}_n} |\alpha_{p,q}|,$$

et $|\alpha_{p,q}| = \frac{|a|^p}{p!} \frac{|b|^q}{q!}$. On en tire la conclusion suivante : si on avait démontré le théorème pour $|a|$ et $|b|$ à la place de a et b, alors on saurait que le membre de droite de cette dernière inégalité tend vers 0, donc le membre de gauche aussi. Ainsi, il est suffisant de montrer le théorème pour tous les nombres réels ≥ 0, l'argument ci-dessus montre qu'il est alors vrai pour tous les complexes.

Nous poursuivons donc en supposant que $a \geq 0$ et $b \geq 0$. Il suffit alors d'observer que $\mathrm{C}_n \subset \mathrm{T}_{2n} \subset \mathrm{C}_{2n}$ pour en déduire que

$$v_n \leq u_{2n} \leq v_{2n}.$$

Ces trois suites sont convergentes, il est donc clair qu'elles ont la même limite. □

Le résultat suivant va être utile pour calculer des dérivées.

Proposition B.3 – *Pour tout nombre complexe z_0, on a*

$$\lim_{z\to z_0} \frac{e^z - e^{z_0}}{z - z_0} = e^{z_0}.$$

C'est bien une limite dans les complexes, pas seulement dans $\mathbb{R}$. Donc à strictement parler ce n'est pas un nombre dérivé.

On dit parfois que la fonction $z \mapsto e^z$ est « holomorphe ».

Démonstration. Prenons d'abord $z_0 = 0$. On écrit

$$\begin{aligned} \frac{e^z - 1}{z} &= 1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots + \frac{z^n}{(n+1)!} + \cdots \\ &= 1 + z\left(\frac{1}{2!} + \frac{z}{3!} + \cdots + \frac{z^n}{(n+2)!} + \cdots \right). \end{aligned}$$

Lorsque $|z| \leq 1$, on a la majoration

$$\begin{aligned} \left| z\left(\frac{1}{2!} + \frac{z}{3!} + \cdots + \frac{z^n}{(n+2)!} + \cdots \right) \right| &\leq |z| \left| \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} + \cdots \right| \\ &= |z|(e-1) \xrightarrow[z\to 0]{} 0. \end{aligned}$$

On a donc bien $\frac{e^z - 1}{z} \to 1 = e^0$ lorsque $z \to 0$.

Maintenant si z_0 est quelconque, on pose $h = z - z_0$ et d'après le théorème précédent,

$$\frac{e^z - e^{z_0}}{z - z_0} = \frac{e^{z_0+h} - e^{z_0}}{h} = e^{z_0} \left(\frac{e^h - 1}{h} \right) \longrightarrow e^{z_0},$$

en utilisant le cas particulier déjà traité. □

Lorsque $x \in \mathbb{R}$, il est clair que $e^x \in \mathbb{R}$. Nous allons nous tourner vers l'étude de $x \mapsto e^x$, vue comme une fonction $\mathbb{R} \to \mathbb{R}$, et bien sûr l'un de nos objectifs est de vérifier qu'il s'agit bien de la fonction abordée en terminale.

Voici de quoi s'en convaincre :

Lemme B.4 – *La fonction* $\exp : \mathbb{R} \to \mathbb{R}$ *est dérivable, et* $\exp' = \exp$. *De plus, il s'agit de l'unique fonction ayant cette propriété et prenant la valeur* 1 *en* 0.

Le mot-clef est peut-être « unique » : au lycée on vous a certainement présenté l'exponentielle comme une fonction égale à sa dérivée, et telle que $\exp(0) = 1$, sans pouvoir *démontrer* son existence. D'après le lemme, la fonction exponentielle que nous avons présentée est la bonne.

Démonstration. Le fait que $\exp' = \exp$ découle directement de la proposition B.3.

Soit maintenant une fonction $f : \mathbb{R} \to \mathbb{R}$ telle que $f' = f$ et $f(0) = 1$. On considère la fonction g définie par $g(x) = e^{-x} f(x)$. On a alors $g'(x) = -e^{-x} f(x) + e^{-x} f'(x) = 0$ puisque $f'(x) = f(x)$. Par suite, la fonction g est constante, disons $g(x) = c$.

Mais alors $f(x) = e^x g(x)$ puisque $e^x e^{-x} = e^0 = 1$, donc $f(x) = ce^x$. Si nous prenons en compte $f(0) = 1$, on en déduit $c = 1$, et finalement $f(x) = e^x$. □

Proposition B.5 – *La fonction exponentielle réalise une bijection* $\mathbb{R} \to \mathbb{R}^{>0}$. *Sa réciproque, que l'on appelle le logarithme népérien et que l'on note* $\ln : \mathbb{R}^{>0} \to \mathbb{R}$, *est dérivable. De plus, on a* $\ln'(x) = \frac{1}{x}$.

Démonstration. On a vu que l'exponentielle ne s'annulait pas. D'après le théorème des valeurs intermédiaires, elle ne peut pas changer de signe sur $\mathbb{R}$, et comme $e^0 = 1$, on conclut que $e^x > 0$ pour $x \in \mathbb{R}$.

Comme $\exp' = \exp > 0$, l'exponentielle est croissante. On a donc $e^x \geq 1$ pour $x \geq 0$. En considérant $g(x) = e^x - x - 1$, qui satisfait $g(0) = 0$ et $g'(x) = e^x - 1 \geq 0$ pour $x \geq 0$, on s'aperçoit que g est croissante et donc reste positive pour $x \geq 0$. En d'autres termes $e^x \geq x + 1$ pour $x \geq 0$, et en particulier

$$\lim_{x \to +\infty} e^x = +\infty .$$

On en déduit

$$\lim_{x \to -\infty} e^x = \lim_{x \to +\infty} e^{-x} = \lim_{x \to +\infty} \frac{1}{e^x} = 0 .$$

D'après le théorème des valeurs intermédiaires, l'image de $\mathbb{R}$ par la fonction exponentielle est un intervalle, contenu dans $]0, +\infty[$. Vues les limites que nous venons de calculer, cette image doit être $]0, +\infty[$ tout entier.

On a donc montré que l'exponentielle était une bijection comme annoncé. Les assertions sur la réciproque se montrent avec la proposition 6.20, et à vrai dire, nous l'avons déjà fait dans l'exemple 6.22. □

Voici une dernière propriété qui était familière au lycée.

Proposition B.6 – *Soit* N *un entier. Alors*

$$\lim_{x \to +\infty} \frac{e^x}{x^{\mathrm{N}}} = +\infty ,$$

et

$$\lim_{x \to -\infty} x^{\mathrm{N}} e^x = 0 .$$

Enfin

$$\lim_{t \to 0} t \ln(t) = 0 .$$

Démonstration. Posons $\mathrm{P}_n(x) = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!}$, de sorte que $e^x = \lim_n \mathrm{P}_n(x)$. Pour $x \geq 0$, on a $e^x \geq \mathrm{P}_n(x)$ pour tout n ; or pour $n > \mathrm{N}$ on a déjà

$$\lim_{x \to +\infty} \frac{\mathrm{P}_n(x)}{x^{\mathrm{N}}} = +\infty ,$$

pour des raisons de degrés. La deuxième limite se calcule en prenant l'inverse, puisque $e^{-x} = \frac{1}{e^x}$. Pour la troisième, on pose $x = x(t) = \ln(t)$, de sorte que $t \ln(t) = xe^x$ avec $x(t) \to -\infty$ lorsque $t \to 0$, d'où le résultat. □

Nous allons maintenant étudier la fonction γ définie sur $\mathbb{R}$ par $\gamma(t) = e^{it}$. En identifiant $\mathbb{C}$ et $\mathbb{R}^2$, on pense à γ comme à une courbe dans le plan. Commençons par une remarque.

LEMME B.7 – *Soit $z \in \mathbb{C}$. Alors $\overline{e^z} = e^{\overline{z}}$.*

Démonstration. Il faut noter que si (u_n) est une suite de complexes telle que $u_n = a_n + ib_n \to \ell = \ell_1 + i\ell_2$, alors $\overline{u_n} = a_n - ib_n \to \overline{\ell} = \ell_1 - i\ell_2$ (puisque $a_n \to \ell_1$ et $b_n \to \ell_2$).

On obtient le lemme en appliquant cette remarque à

$$u_n = \sum_{k=0}^{n} \frac{z^n}{n!},$$

qui converge vers e^z, alors que

$$\overline{u_n} = \sum_{k=0}^{n} \frac{\overline{z}^n}{n!}$$

converge vers $e^{\overline{z}}$. □

Puisque $\overline{it} = -it$ quand $t \in \mathbb{R}$, nous avons $\overline{e^{it}} = e^{-it}$, et donc

$$|e^{it}|^2 = e^{it}\,\overline{e^{it}} = e^0 = 1.$$

Par suite $|\gamma(t)| = 1$: la courbe γ prend ses valeurs dans

$$\mathrm{C} = \{z \in \mathbb{C} \text{ tels que } |z| = 1\},$$

c'est-à-dire le cercle de centre $(0,0)$ et de rayon 1. L'un de nos grands objectifs est de montrer que γ passe par chaque point du cercle C.

LEMME B.8 – *La fonction γ est dérivable et $\gamma'(t) = ie^{it}$.*

Démonstration. On calcule

$$\frac{e^{it} - e^{it_0}}{t - t_0} = i\frac{e^{it} - e^{it_0}}{it - it_0} \longrightarrow ie^{it_0},$$

d'après la proposition B.3 appliquée en $z_0 = it_0$. □

On a notamment $|\gamma'(t)| = |ie^{it}| = 1$. En termes plus géométriques : le point $\gamma(t)$ se déplace *à vitesse constante* le long du cercle. Intuitivement, c'est peut-être déjà suffisant pour se convaincre que γ va faire tout le tour de C.

DÉFINITION B.9 – On définit une fonction cos sur $\mathbb{R}$ par la formule

$$\cos(t) = \mathfrak{Re}(e^{it}) = \frac{e^{it} + e^{-it}}{2}.$$

On l'appelle le *cosinus*. De même on définit une fonction sin sur $\mathbb{R}$ par la formule

$$\sin(t) = \mathfrak{Im}(e^{it}) = \frac{e^{it} - e^{-it}}{2i},$$

et on l'appelle le *sinus*.

On a donc $e^{it} = \cos(t) + i\sin(t)$. De plus la relation $|e^{it}|^2 = 1$ donne $\cos(t)^2 + \sin(t)^2 = 1$.

À partir des définitions et du lemme précédent, on obtient tout de suite :

LEMME B.10 – *La fonction* cos *est dérivable et* $\cos' = -\sin$. *De même la fonction* sin *est dérivable et* $\sin' = \cos$.

Voici un résultat fondamental.

THÉORÈME B.11 – *Il existe un unique nombre réel positif, noté π, tel que*

$$e^{it} = 1 \iff t = 2n\pi \text{ avec } n \in \mathbb{Z}.$$

De plus, on a $e^{i\pi} + 1 = 0$.

Démonstration. Commençons par une petite étude du sinus au voisinage de 0. Puisque $e^0 = 1$, on a $\sin(0) = 0$ et $\sin'(0) = \cos(0) = 1$. Comme la fonction cosinus est continue (et même dérivable), la proposition 4.19 nous assure qu'il existe un intervalle $]a,b[$ avec $a < 0 < b$ tel que $\cos(t) > 0$ pour $t \in]a,b[$. La fonction sinus est donc strictement croissante sur cet intervalle, et en particulier $\sin(t) \neq 0$ si $t \neq 0$ et $t \in]a,b[$; on en déduit également que $e^{it} \neq 1$ pour $t \neq 0$ et $t \in]a,b[$.

Montrons maintenant que la fonction cosinus peut s'annuler. Procédons par l'absurde : si $\cos(t) \neq 0$ pour tout t, alors on aurait $\cos(t) > 0$ d'après le théorème des valeurs intermédiaires, et sin serait croissante sur tout $\mathbb{R}$. En particulier, on aurait $\sin(t) \geq \sin(b) > 0$ pour $t \geq b$. Mais alors, regardons la fonction g définie par $g(t) = \cos(t) + \sin(b)t - 1$ qui vérifie $g(0) = 0$ et $g'(t) = -\sin(t) + \sin(b) \leq 0$. Elle est donc décroissante, et par suite $g(t) \leq 0$ pour $t \geq 0$, ce qui donne $\cos(t) \leq 1 - \sin(b)t$. C'est absurde, puisqu'on en déduit que $\cos(t) \to -\infty$ lorsque $t \to +\infty$, alors que bien sûr le cosinus prend ses valeurs dans $[-1,1]$ à cause de la relation $\cos(t)^2 + \sin(t)^2 = 1$.

On peut donc trouver t_0 tel que $\cos(t_0) = 0$, et on peut même s'arranger pour que $t_0 > 0$. On a alors $\sin(t_0)^2 = 1$ donc $\sin(t_0) = \pm 1$ et $e^{it_0} = \pm i$. Comme $i^4 = (-i)^4 = 1$, on a $e^{4it_0} = (e^{it_0})^4 = 1$.

Passons au théorème proprement dit. Posons

$$\mathrm{K} = \{t \in \mathbb{R} \mid e^{it} = 1\},$$

et

$$\mathrm{A} = \{t \in \mathrm{K} \text{ et } t > 0\}.$$

Nous avons prouvé que $\mathrm{A} \neq \emptyset$, puisque cet ensemble contient l'élément $4t_0 > 0$. Nous avons également prouvé que A ne contient aucun élément dans l'intervalle $]a,b[$. On peut donc poser $\ell = \inf \mathrm{A}$, ce nombre est alors bien défini et $\ell \geq b > 0$. Enfin, on pose $\pi = \frac{\ell}{2}$ (c'est la définition du nombre π).

Il faut commencer par vérifier que $2\pi = \ell$ appartient à K (et même à A). En effet, par définition de l'inf il existe une suite (t_n) qui converge vers ℓ avec $t_n \in \mathrm{A}$, donc $e^{it_n} = \gamma(t_n) = 1$. Par continuité de γ, on a $\gamma(t_n) \to \gamma(\ell)$, et donc $\gamma(\ell) = 1$, ce qui signifie bien que $\ell \in \mathrm{K}$.

Soit maintenant $t \in \mathrm{K}$ quelconque, et soit n l'unique nombre entier tel que

$$n \leq \frac{t}{2\pi} < n + 1,$$

de sorte que $0 \leq t - 2n\pi < 2\pi$. On note que

$$e^{i(t-2n\pi)} = e^{it}e^{-2ni\pi} = e^{it}(e^{2i\pi})^{-n} = 1 \times 1^{-n} = 1.$$

Ainsi $t - 2n\pi \in \mathrm{K}$, mais par définition de $2\pi = \ell = \inf \mathrm{A}$, le seul élément de K dans $[0, 2\pi[$ est 0. On en conclut bien que $t = 2n\pi$, comme on souhaitait le montrer.

Enfin, le nombre $e^{i\pi}$ vérifie $(e^{i\pi})^2 = e^{2i\pi} = 1$, donc $e^{i\pi} = \pm 1$. Mais $\pi < 2\pi$ donc $e^{i\pi} = 1$ est exclu. Finalement on a bien $e^{i\pi} = -1$, ce qui achève la démonstration. □

PROPOSITION B.12 – *Sur l'intervalle $[0, \frac{\pi}{2}]$, la fonction cosinus est décroissante. Son image est l'intervalle $[0,1]$ entier.*

Sur le même intervalle, la fonction sinus est croissante. Son image est également l'intervalle $[0,1]$ entier.

Démonstration. Le nombre $x = e^{i\frac{\pi}{2}}$ vérifie $x^2 = e^{i\pi} = -1$, donc $x = \pm i$. On a donc $\cos(\frac{\pi}{2}) = 0$ et $\sin(\frac{\pi}{2}) = \pm 1$. (Dans un instant nous allons trouver le bon signe.)

Le nombre $\frac{\pi}{2}$ est le plus petit nombre réel positif pour lequel le cosinus s'annule : en effet si t_0 est un tel nombre, nous avons vu au cours de la démonstration du théorème que $4t_0$ vérifie $e^{4it_0} = 1$; nous savons alors que $4t_0$ est un multiple de 2π, donc t_0 est un multiple de $\frac{\pi}{2}$. Ainsi la fonction cosinus ne s'annule pas sur $[0, \frac{\pi}{2}[$, et donc elle ne change pas de signe (valeurs intermédiaires). Comme $\cos(0) = 1$, on en déduit $\cos(t) \geq 0$ sur cet intervalle.

La fonction sinus est donc croissante sur le même intervalle puisque $\sin' = \cos$, et puisque $\sin(0) = 0$, on a $\sin(t) \geq 0$ pour $t \in [0, \frac{\pi}{2}]$. En particulier $\sin(\frac{\pi}{2}) = +1$.

Enfin, la relation $\cos' = -\sin$ montre que la fonction cosinus est décroissante, toujours sur le même intervalle.

Pour conclure, c'est le théorème des valeurs intermédiaires qui garantit que les deux fonctions en question prennent bien toutes les valeurs entre 0 et 1. □

Nous avons atteint notre objectif :

PROPOSITION B.13 – *Tout point du cercle unité C est de la forme e^{it} pour au moins un nombre $t \in \mathbb{R}$. De plus, deux nombres t et u vérifient $e^{it} = e^{iu}$ si et seulement si $t = u + 2n\pi$ avec $n \in \mathbb{Z}$.*

Démonstration. Si $e^{it} = e^{iu}$, alors $e^{i(t-u)} = 1$, donc la dernière partie de l'énoncé est une conséquence du théorème.

Soit maintenent $x + iy$ un nombre complexe de module 1, c'est-à-dire que $x^2 + y^2 = 1$. Dans un premier temps, supposons que $x \geq 0$ et $y \geq 0$, de sorte que $y = \sqrt{1 - x^2}$. Le nombre x appartient à $[0,1]$, il est donc de la forme $x = \cos(t)$ pour un unique nombre $t \in [0, \frac{\pi}{2}]$, d'après la proposition précédente. On a alors $\sin(t) \geq 0$ et bien sûr $\cos(t)^2 + \sin(t)^2 = 1$, donc $\sin(t) = \sqrt{1 - \cos(t)^2} = \sqrt{1 - x^2} = y$. Finalement $e^{it} = x + iy$, ce qui prouve le résultat dans ce cas.

Si $y \leq 0$, on prend d'abord un t tel que $e^{it} = x - iy$ (cas précédent), et alors $e^{-it} = x + iy$ (lemme B.7). Et enfin, si $x \leq 0$, on prend t tel que $e^{it} = -x - iy$, et alors $e^{i(\pi+t)} = -e^{it} = x + iy$. □

exercice : donnez une démonstration de toutes les propriétés des fonctions sinus et cosinus que vous connaissez depuis quelques

PROPOSITION B.14 – *Tout nombre complexe non-nul $z \in \mathbb{C}^*$ peut s'écrire $z = \rho e^{i\theta}$ avec $\rho > 0$ et $\theta \in \mathbb{R}$. Le nombre ρ est unique, et en fait $\rho = |z|$; le nombre θ est déterminé à un multiple de 2π près.*

En conséquence, tout nombre complexe non-nul z peut s'écrire $z = e^w$ pour $w \in \mathbb{C}$.

Démonstration. Il suffit de considérer $\frac{z}{|z|}$: c'est un complexe de module 1, donc $\frac{z}{|z|} = e^{i\theta}$ d'après la proposition B.13. Ainsi $z = |z|e^{i\theta}$. La même proposition indique que θ est déterminé à un multiple de 2π près.

En prenant $x = \ln(|z|)$, et $w = x + i\theta$, nous avons bien $e^w = e^x e^{i\theta} = |z|e^{i\theta} = z$. □

Voici une application importante :

COROLLAIRE B.15 – *Soit $n \geq 1$ un entier. Tout nombre complexe z possède des racines n-ièmes.*

Plus précisément, si $z \neq 0$, il existe exactement n nombres (distincts) $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ tels que $\alpha_k^n = z$.

La démonstration va indiquer comment trouver explicitement ces racines.

Démonstration. On écrit $z = \rho e^{i\theta}$ et l'on prend $\alpha = \sqrt[n]{\rho} e^{i\frac{\theta}{n}}$; on a alors effectivement $\alpha^n = \rho e^{i\theta} = z$.

Voyons les autres possibilités. Si $\beta = re^{i\phi}$ vérifie $\beta^n = 1 = r^n e^{ni\phi}$, alors on doit avoir $r^n = \rho$ et $n\phi = \theta + 2k\pi$ avec $k \in \mathbb{Z}$. Comme $r \geq 0$, on en déduit $r = \sqrt[n]{\rho}$. D'autre part

$$\phi = \frac{\theta}{n} + \frac{2k\pi}{n},$$

donc

$$e^{i\phi} = e^{i\frac{\theta}{n}} e^{\frac{2ki\pi}{n}},$$

de sorte que finalement $\beta = \alpha e^{\frac{2ki\pi}{n}}$ pour un certain $k \in \mathbb{Z}$. Réciproquement si β est de cette forme alors $\beta^n = \alpha^n = z$, puisque $\left(e^{\frac{2ki\pi}{n}}\right)^n = 1$.

voir les exercices 42, 43, 44, 45, 2939

Si k et ℓ sont deux entiers, alors les nombres $e^{\frac{2ki\pi}{n}}$ et $e^{\frac{2\ell i\pi}{n}}$ sont égaux précisément lorsqu'il existe un entier m tel que $\ell = k + mn$ (autrement dit, lorsque k et ℓ sont égaux modulo n). On en déduit que pour $0 \leq k \leq n-1$, les nombres $e^{\frac{2ki\pi}{n}}$ sont distincts, et que tout nombre de la forme $e^{\frac{2\ell i\pi}{n}}$ avec ℓ entier est dans cette liste. Ce sont les n racines « de l'unité ».

Quel lien voyez-vous entre les racines n-ièmes et $\mathbb{Z}/n\mathbb{Z}$?

Finalement on a bien n nombres qui conviennent, à savoir $\alpha_k = \alpha e^{\frac{2ki\pi}{n}}$ pour $0 \leq k \leq n-1$. □

Ce théorème est cité dans le premier chapitre d'algèbre (sous le numéro 11.9) sans démonstration. Revoici l'énoncé :

Théorème B.16 – *Tout polynôme de degré* ≥ 1 *dans* $\mathbb{C}[X]$ *possède une racine dans* $\mathbb{C}$.

Démonstration. Soit donc $P \in \mathbb{C}[X]$ de degré ≥ 1 ; écrivons

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n,$$

avec $a_n \neq 0$.

La fonction définie sur $\mathbb{C}$ par $z \mapsto P(z)$ est continue. De plus en écrivant

$$P(z) = a_n z^n \left(1 + \frac{a_{n-1}}{a_n}\frac{1}{z} + \frac{a_{n-2}}{a_n}\frac{1}{z^2} + \cdots + \frac{a_0}{a_n}\frac{1}{z^n}\right),$$

on voit que pour toute suite $(z_n)_{n\geq 0}$ de nombres complexes tels que $|z_n| \to +\infty$, on a également $|P(z_n)| \to +\infty$. La fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(z) = |P(z)|$ satisfait donc les hypothèses de la proposition 5.8, et on conclut qu'il existe $z_0 \in \mathbb{C}$ tel que $f(z) \geq f(z_0)$ pour tout $z \in \mathbb{C}$. Nous allons montrer que $f(z_0) = 0$, donc $P(z_0) = 0$, ce qui établit le théorème.

Procédons par l'absurde, et supposons que $P(z_0) \neq 0$. Pour se faciliter les choses, considérons

$$Q = \frac{P(z_0 + X)}{P(z_0)};$$

alors Q est un polynôme de même degré que P, la quantité $|Q(z)|$ atteint un minimum en 0, et ce minimum vaut $Q(0) = 1$. Écrivons

$$Q(z) = 1 + b_d z^d + b_{d+1} z^{d+1} + \cdots + b_n z^n,$$

avec $b_d \neq 0$, et $d \geq 1$. Maintenant, pour se débarrasser de b_d, passons au polynôme

$$R(X) = Q(\alpha X),$$

où α vérifie $\alpha^d = -\frac{1}{b_d}$: un tel nombre existe d'après le corollaire B.15. Ce R est un polynôme de degré n, la quantité $|R(z)|$ atteint un minimum en 0, et ce minimum vaut encore $R(0) = 1$; mais de plus

$$R(z) = 1 - z^d + c_{d+1} z^{d+1} + \cdots + c_n z^n,$$

pour des coefficients c_i dont la valeur n'a pas d'importance pour la suite.

Montrons que la relation $|R(z)| \geq 1$ nous mène à une contradiction. Prenons $x \in \mathbb{R}$ et $x > 0$, et écrivons $R(x) = 1 - x^d + x^d \varepsilon(x)$, avec $\varepsilon(x) \to 0$ lorsque $x \to 0$. Avec l'inégalité triangulaire il vient

$$\begin{aligned} 1 &\leq |1 - x^d| + |x^d \varepsilon(x)| \\ &= 1 - x^d + x^d |\varepsilon(x)|. \end{aligned}$$

Pour x suffisamment petit on aura $|\varepsilon(x)| \leq 1/2$, et en arrangeant les termes de l'inégalité ci-dessus on aura $2 \leq 1$, une contradiction. □

– Fin de la première lecture –

DÉFINITION B.17 – Soit $A = (a_{ij}) \in M_{n,m}(\mathbb{R})$. Sa *norme* (euclidienne) est

$$\|A\| = \left(\sum_{i,j} a_{ij}^2\right)^{\frac{1}{2}}.$$

De même, soit $A = (a_{ij}) \in M_{n,m}(\mathbb{C})$. Sa norme euclidienne est

$$\|A\| = \left(\sum_{i,j} |a_{ij}|^2\right)^{\frac{1}{2}}.$$

Quelques remarques s'imposent. Une matrice réelle dont la taille est $n \times m$ est constituée de nm coefficients, et on peut identifier l'ensemble $M_{n,m}(\mathbb{R})$ avec $\mathbb{R}^{nm}$. Ceci étant fait, la norme d'une matrice n'est autre que la norme euclidienne du vecteur correspondant de $\mathbb{R}^{nm}$, que nous connaissons bien (définition 3.28). De même avec les matrices complexes, on identifie $M_{n,m}(\mathbb{C})$ avec $\mathbb{C}^{nm}$; de plus on peut identifier $\mathbb{C}$ avec $\mathbb{R}^2$, de telle sorte que $z = x + iy$ correspond à (x, y), et alors $|z|^2 = x^2 + y^2$; par suite $\mathbb{C}^{nm}$ peut être vu comme $\mathbb{R}^{2nm}$, et au total la norme d'une matrice complexe n'est rien d'autre que la norme du vecteur de $\mathbb{R}^{2nm}$ correspondant.

Pour les calculs, il va être utile de faire la remarque suivante. Si les colonnes de A sont $C_1, C_2, \ldots, C_m$, alors

$$\|A\|^2 = \|C_1\|^2 + \cdots + \|C_m\|^2 ;$$

de même si les lignes de A sont $L_1, \ldots, L_n$, alors

$$\|A\|^2 = \|L_1\|^2 + \cdots + \|L_n\|^2 .$$

La différence entre les matrices et les vecteurs, évidemment, est que l'on peut multiplier les matrices. Le résultat suivant donne le lien entre les normes et la multiplication.

LEMME B.18 – *Soient* $A \in M_{n,m}(\mathbb{C})$ *et* $B \in M_{m,\ell}(\mathbb{C})$. *Alors*

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

Démonstration. Écrivons $A = (a_{ij})$ et notons $L_1, \ldots, L_n$ les lignes de A ; de même écrivons $B = (b_{ij})$ et notons $C_1, \ldots, C_\ell$ les colonnes de B. La matrice AB possède, sur sa ligne i et dans la colonne j, le coefficient

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} .$$

On a donc les inégalités

$$\begin{aligned} |c_{ij}| &\leq \sum_{k=1}^{m} |a_{ik}||b_{kj}| \\ &\leq \left(\sum_k |a_{ik}|^2\right)^{\frac{1}{2}} \left(\sum_k |b_{kj}|^2\right)^{\frac{1}{2}} \\ &= \|L_i\| \cdot \|C_j\|. \end{aligned}$$

La première est l'inégalité triangulaire, la deuxième est l'inégalité de Cauchy-Schwarz (lemme 3.32). En prenant la somme, il vient

$$\begin{aligned} \|AB\|^2 = \sum_{i,j} |c_{ij}|^2 &\leq \sum_{i,j} \|L_i\|^2 \cdot \|C_j\|^2 \\ &= \left(\sum_i \|L_i\|^2\right)\left(\sum_j \|C_j\|^2\right) \\ &= \|A\|^2 \cdot \|B\|^2 . \end{aligned}$$

Le résultat en découle. □

Lemme B.19 – *Soit* A *une matrice carrée. On note*

$$S_n(A) = \sum_{k=0}^{n} \frac{1}{k!} A^k.$$

Alors la suite $(S_n(A))_{n\geq 0}$ *possède une limite.*

Avant de donner la démonstration, indiquons tout de suite :

Définition B.20 – Soit A une matrice carrée, à coefficients complexes. Son *exponentielle* est

$$\exp(A) = e^A = \lim_{n\to\infty} S_n(A) = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k.$$

Démonstration. On utilise la convergence absolue, c'est-à-dire le théorème 3.30, qui nous assure qu'il suffit de vérifier que

$$\lim_{n\to\infty} \sum_{k=0}^{n} \left\| \frac{1}{k!} A^k \right\|$$

existe. Or, on a

$$\left\| \frac{1}{k!} A^k \right\| \leq \frac{1}{k!} \|A\|^k,$$

d'après le lemme B.18. Si on note

$$u_n = \sum_{k=0}^{n} \left\| \frac{1}{k!} A^k \right\|,$$

on a donc

$$u_n \leq \sum_{k=0}^{n} \frac{1}{k!} \|A\|^k \leq \sum_{k=0}^{+\infty} \frac{1}{k!} \|A\|^k = e^{\|A\|}.$$

La suite (u_n) est donc croissante et majorée, et on en conclut qu'elle possède bien une limite. □

Pour l'instant, on ne sait calculer que des exemples très simples :

Exemple B.21 – Prenons une matrice diagonale :

$$A = \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}.$$

On a donc

$$A^n = \begin{pmatrix} x^n & 0 \\ 0 & y^n \end{pmatrix},$$

et

$$S_n(A) = \begin{pmatrix} \sum_{k=0}^{n} \frac{x^k}{k!} & 0 \\ 0 & \sum_{k=0}^{n} \frac{y^k}{k!} \end{pmatrix}.$$

En passant à la limite, on obtient :

$$e^A = \begin{pmatrix} e^x & 0 \\ 0 & e^y \end{pmatrix}.$$

Proposition B.22 – *L'exponentielle de matrice possède les propriétés suivantes :*

1. *Si* P *est inversible, alors*
$$e^{P^{-1}AP} = P^{-1} e^A P.$$
2. *Si* AB = BA, *alors* $e^A B = B e^A$.
3. *Si* AB = BA, *alors*
$$e^{A+B} = e^A e^B.$$

La propriété (3) est la plus importante, bien sûr. Attention, l'hypothèse AB = BA est nécessaire !

Démonstration. Pour le (1), on calcule d'abord

$$(P^{-1}AP)^2 = P^{-1}APP^{-1}AP = P^{-1}A^2P,$$

et de même

$$(P^{-1}AP)^3 = P^{-1}AP(P^{-1}AP)^2 = P^{-1}APP^{-1}A^2P = P^{-1}A^3P.$$

Par récurrence on obtient pour tout n :

$$(P^{-1}AP)^n = P^{-1}A^nP.$$

On voit que $S_n(P^{-1}AP) = P^{-1}S_n(A)P$, d'où la formule (1) en passant à la limite.

Pour le (2), on passe à la limite dans la relation évidente $S_n(A)B = BS_n(A)$ (ou alors, si B est inversible, on utilise la relation $B^{-1}AB = A$, d'où $B^{-1}e^AB = e^A$ d'après le (1)).

Pour le (3), c'est essentiellement la même démonstration que pour le théorème B.2. L'hypothèse AB = BA est utilisée pour pouvoir utiliser la formule de Newton qui donne le développement de $(A+B)^n$ (elle est fausse si $AB \neq BA$, par exemple $(A+B)^2 = A^2 + AB + BA + B^2$ en général). On constate (toujours en suivant les étapes de la démonstration dans le cas des complexes) qu'il suffit de montrer la formule pour $\|A\|$ et $\|B\|$ au lieu de A et B, et nous savons que l'identité est vraie pour les nombres réels. □

Exemple B.23 – Prenons

$$A = \begin{pmatrix} 11 & 18 \\ -6 & -10 \end{pmatrix}.$$

Pour calculer e^A, on essaie de la mettre sous la forme $P^{-1}BP$ où B est le plus simple possible. Dans le chapitre « Diagonalisation », nous verrons de nombreuses techniques pour faire ça ; pour l'instant, admettons que l'on nous ait soufflé que, en posant

$$P = \begin{pmatrix} -3 & -2 \\ 2 & 1 \end{pmatrix},$$

alors

$$P^{-1}AP = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} = D.$$

On a donc

$$e^A = e^{PDP^{-1}} = Pe^DP^{-1},$$

d'après le (1) de la proposition. On peut calculer e^D sans peine comme dans l'exemple précédent, et finalement

$$e^A = \begin{pmatrix} -3 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} e^{-1} & 0 \\ 0 & e^2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & -3 \end{pmatrix} = \begin{pmatrix} -3e^{-1} + 4e^2 & -6e^{-1} + 6e^2 \\ 2e^{-1} - 2e^2 & 4e^{-1} - 3e^2 \end{pmatrix}.$$

Voici comment exploiter la propriété (3). Prenons

$$A = \begin{pmatrix} 5 & 7 \\ 0 & 5 \end{pmatrix}.$$

On pose alors A = D + N avec

$$D = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \quad \text{et} \quad N = \begin{pmatrix} 0 & 7 \\ 0 & 0 \end{pmatrix}.$$

On vérifie que DN = ND, donc $e^A = e^De^N$. De nouveau, on peut calculer e^D facilement puisque la matrice est diagonale. Pour N, on constate que $N^2 = 0$, ce qui ramène son exponentielle à $e^N = Id + N$. Finalement

$$e^A = e^D(Id + N) = \begin{pmatrix} e^5 & 7e^5 \\ 0 & e^5 \end{pmatrix}.$$

voir l'exercice 1716

Proposition B.24 – *Soit* A *une matrice complexe, de taille* $n \times n$. *La fonction* $\gamma\colon \mathbb{R} \to \mathrm{M}_n(\mathbb{C})$ *définie par* $\gamma(t) = e^{t\mathrm{A}}$ *est dérivable, et sa dérivée est* $\gamma'(t) = \mathrm{A}e^{t\mathrm{A}} = \mathrm{A}\gamma(t)$.

De plus, cette propriété caractérise γ, *c'est-à-dire que si* $c\colon \mathbb{R} \to \mathrm{M}_n(\mathbb{C})$ *vérifie* $c'(t) = \mathrm{A}c(t)$ *et* $c(0) = \mathrm{Id}$, *alors on a* $c(t) = \gamma(t)$ *pour tout* t.

Il faut noter que, si s et t sont deux nombres réels, alors les matrices $s\mathrm{A}$ et $t\mathrm{A}$ commutent, donc $e^{(s+t)\mathrm{A}} = e^{s\mathrm{A}}e^{t\mathrm{A}}$. En d'autres termes $\gamma(s+t) = \gamma(s)\gamma(t)$.

Démonstration. Le calcul de $\gamma'(t)$ est tout à fait analogue à la démonstration de la proposition B.3.

Montrons l'unicité. Supposons donc que $c'(t) = \mathrm{A}c(t)$ et $c(0) = \mathrm{I}d$, et définissons

$$f(t) = e^{-t\mathrm{A}}c(t).$$

D'après le lemme 6.25, la dérivée de f est donnée par

$$f'(t) = (-\mathrm{A})e^{-t\mathrm{A}}c(t) + e^{-t\mathrm{A}}c'(t) = e^{-t\mathrm{A}}\left[-\mathrm{A}c(t) + c'(t)\right] = 0.$$

On a utilisé le fait que $\mathrm{A}e^{-t\mathrm{A}} = e^{-t\mathrm{A}}\mathrm{A}$, d'après le (2) de la proposition B.22. Puisque $f'(t) = 0$, on en conclut que f est constante, donc pour tout t on a $f(t) = f(0) = \mathrm{Id}$. Ceci donne $e^{-t\mathrm{A}}c(t) = \mathrm{Id}$, et en multipliant par $e^{t\mathrm{A}}$ on en tire bien $c(t) = e^{t\mathrm{A}} = \gamma(t)$. □